

REVIEW

Progress in Activities of WDS-China Data Centers

Juanle Wang^{1,5}, Kun Bu², Yanjie Wang^{1,3} and Yating Shao^{1,4}

¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, CN

² Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, CN

³ College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing, CN

⁴ School of Geomatics and Marine Information, Jiangsu Ocean University, Jiangsu, CN

⁵ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, CN

Corresponding author: Juanle Wang (wangjl@igsrr.ac.cn)

The World Data System (WDS) plays an important role in promoting global scientific data management, exchange, and sharing. There are 8 WDS data centers in mainland China including study areas of astronomy, space science, global change, renewable resources and environmental, cold and dry regions, microbiology, geophysics, and the ocean. This paper summarizes the current status of the WDS China data centers, along with their major progress in recent years. This progress includes a clearinghouse for metadata exchange, research data archival, historical data saving, international data exchange, data publishing models, CoreTrustSeal certification, open repositories for the scientific community, science popularization services, international training workshops, and awards. We conclude with a discussion of the opportunities and challenges faced by WDS China data centers.

Keywords: World Data System; China Data Center; activities; progress; scientific data management

1 Introduction

In this information age, scientific data is integral not only to scientific and technological infrastructure but also to national strategic resources. Scientific Data Centers are one of the important carriers of scientific data management (Wang et al. 2017). The World Data System (WDS) is an Interdisciplinary Body of the International Science Council (ISC; formerly ICSU). The WDS builds on the 50+ year legacy of the World Data Centres and Federation of Astronomical and Geophysical data analysis Services established by ICSU to manage data generated by the International Geophysical Year (1957–1958). In 1988, the following 9 WDCs were established in China by the former ICSU Panel on World Data Centers, including WDC for Astronomy, Beijing, WDC for Space Science, Beijing, WDC for Geophysics, Beijing, WDC for Meteorology, Beijing, WDC for Oceanography, Tianjin, WDC for Glaciology and Geocryology, Lanzhou, WDC for Seismology, Beijing, WDC for Geology, Beijing, and WDC for Renewable Resources and Environment. These centers were reviewed successfully by WDC Panel team in 2005. WDS China has 8 data centers in the mainland at present. This paper summarizes the activities and progress of these centers in recent years.

2 Status and Activities

WDS China Data Centers are organized around the framework of the earth system and solar-terrestrial space as shown in **Figure 1**. Outer space, is WDS China Astronomical Data Center (CASDC) in Beijing, while solar-terrestrial space is covered by the WDS China Space Science Data Center (CSSDC) in Beijing. For the earth system, there are many data centers across the sub-disciplines of atmosphere, biosphere, lithosphere, hydrosphere (cryosphere), and anthroposphere. They are the Global Change Scientific Research Data Publishing System (GCdataPR), Beijing; the World Data Center for Renewable Resources and Environment (WDC-RRE), Beijing; the World Microbiological Data Center (WDCM), Beijing; the WDS for Cold Dry Area Science Data

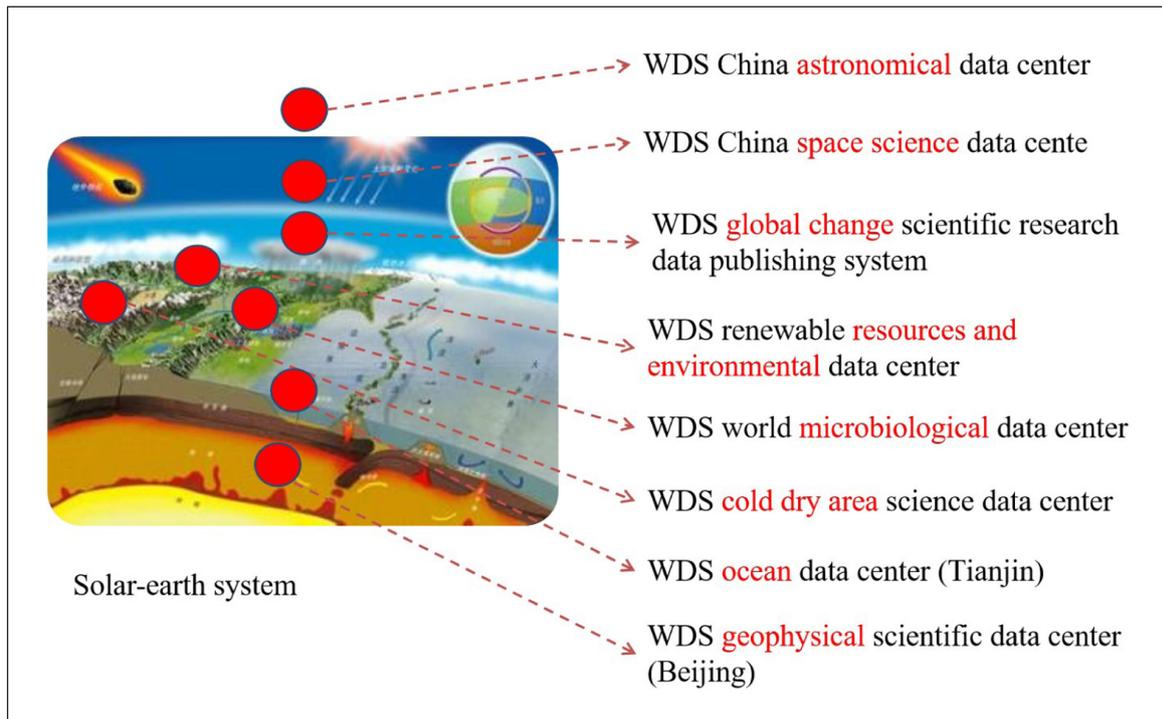


Figure 1: Distribution of WDS China data centers in the earth system and solar-terrestrial space.

Center, Lanzhou; the WDS for Geophysical Scientific Data Center, Beijing; and the WDS for Ocean Data Center in Tianjin.

The CAsDC is supported by the National Astronomical Observatories, CAS (<http://explore.china-vo.org/>). Data received from Guoshoujing's Telescope (Large Sky Area Multi-Object Fiber Spectroscopic Telescope) is archived and released in the CAsDC. As a distribution platform, the CAsDC provides global services and connects all the astronomical observatories in China with the Alibaba global cloud facility. It has been a member of the International Virtual Observatory Alliance since 2002.

The CSSDC is supported by the National Space Science Center, CAS (<http://www.cssdc.ac.cn/>). It integrates and optimizes the data of space science, focusing on the integrity, systematization, and standardization of data management in space science, ensuring the permanent security and long-term availability of space science data and improving the level and efficiency of data application while also exchanging and sharing international space science data.

The GCdataPR is supported by the Institute of Geographic Sciences and Natural Resources Research, CAS and the Geographical Society of China (<http://www.geodoi.ac.cn>). It establishes a set of mechanisms and management methods for data publishing, preservation and sharing in the field of global change scientific research. GCdataPR advocates an innovative data sharing approach that integrates metadata, data products, and data papers.

The WDC-RRE is supported by the Institute of Geographic Sciences and Natural Resources Research, CAS (<http://eng.wdc.cn/>). This repository's principal databases include basic geographic, natural resources, population and social economy, disaster risk reduction, land use and land cover, Loess Plateau agriculture and environment, temperature products, and special region or thematic databases. It is also one of the sub-centers of the National Earth System Science Data Sharing Infrastructure in China.

The WDS for Cold Dry Area Science Data Center is supported by the Cold and Arid Regions Environmental and Engineering Research Institute, CAS (<http://card.westgis.ac.cn>). It mainly includes scientific data on the cryosphere (glacier, snow, frozen soil), deserts and desertification, continental river basins in arid regions, and the critical scientific data set of land surface processes in cold and arid regions. It advocates the publishing of scientific data with unique digital object identification.

The WDCM is supported by the Institute of Microbiolmicrobes and was set up as a data center of the World Federation for Culture Collections, CAS (<http://www.wdcm.org>). It is a vehicle for networking microbial resource centers of various types of WFCC). The WDCM is constructing a data management system and a

global catalog to help organizing, discovering and exploring the data resources of its member collections (Wu et al. 2013).

The WDS for Geophysical Scientific Data Center (Beijing) is supported by the Institute of Geology and Geophysics, CAS (<http://www.geophys.ac.cn>). The basic tasks of this data center are collection, handling and storage of scientific data and providing access to scientific research. It obtains global space environment and solid earth observation data through its observations in China and participation in international joint observation and data exchange projects.

The WDS for Ocean Data Center (Tianjin) is supported by the China Oceanic Information Network (<http://www.cmoc-china.cn>). It is responsible for the management of national marine data and information resources, providing guidance on and scientific management of national marine data and information, along with information and technical support for the marine economy, marine sustainable development, marine environmental protection, public services, and carrying out relevant research.

3 Progress achieved by WDS China Data Centers

1. Clearinghouse for metadata exchange

The framework for WDS China Common Clearinghouse was preliminarily put forward and a prototype system was built using Pycsw, which is a Python realization approach of OGC Catalogue Services for the Web (CSW) standard. The pycsw technical framework was used to establish metadata management systems and allowed its metadata standards to be compatible with other international and national standards. The metadata capture module was built based on data harvesting to making all the metadata information could be accessed among the data centers in China. The initial progress can be seen in the exchange system (the website of WDS China, <http://www.wds-china.org/>). The job has been involved in the WDS Harvestable Metadata Working Group in 2019.

2. Research data archiving

The CSSDC gradually archived and released data of major projects, such as the Space-based multi-band astronomical Variable Objects Monitor (SVOM) Strategic Priority Program on Space Science in CAS, the International Space Weather Meridian Circle Program and others. The WDS for Cold Dry Area Science Data Center has made rapid advances in archiving major science and technology projects in China. Datasets of Heihe Watershed Allied Telemetry Experimental Research were archived and released based on a series of careful quality control procedures throughout sensor calibration, data collection, data processing, and datasets generation.

3. International cooperation data exchange

The WDCM is establishing a Global Catalog of Microorganisms (GCM), which is expected to be a robust, reliable and user-friendly system to help culture collections manage, disseminate and share the information related to their holdings. The GCM includes information on strain, taxonomy, isolation, application, paper, patents, sequence and protein. Up to now, there are 48 countries, 118 institutes and 447695 strains in the GCM. The WDS for Ocean Data Center has established a formal relationship of marine data exchange with over 130 marine institutions in more than 60 countries and is maintaining a close relationship of data exchange with over 30 major national oceanographic data centers. The data center is involved in global collaboration projects that include: the Global Ocean Observing System (GOOS), the Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM), the Global Sea Level Observing System (GLOSS), the Global Temperature and Salinity Profile Plan (GTSP), the Array for Real-time Geostrophic Oceanography (ARGO), the North Pacific Marine Science Organization (PICES), and more. It also records 485 tide prediction sites around the world. CSSDC cooperated with the European Space Agency and French Space Agency in missions of Solar wind Magnetosphere Ionosphere Link Explorer (SMILE) and SVOM. During the past two years, the CSSDC was in charge of the construction of the joint observational network of China and Brazil supported by CAS.

4. Historical data saving

Historical ionospheric data collected by the WDS for the Geophysical Scientific Data Center (Beijing) comes from decades of manually created records, including around 50 years of photographic film records, and about 20 years of digital records. The ionospheric characteristic parameter database covers the continuous observation data of more than one solar activity cycle (11 years) in and around China, especially the continu-

ous observation data of 70 years in Wuhan. This is the ionospheric characteristic parameter observation data with the longest observation in China.

5. Data publishing model

GCdataPR is becoming known as the new pattern of demonstration data centers of the world (Ma et al. 2019). It advocates an innovative data sharing approach that integrates metadata, data products, and data papers. Its published data set statistics is shown in **Table 1** since June 2014. It has become the open repository of 59 academic journals published by Chinese and American institutes.

6. Open repository

Since August 1, 2019, the American Geophysical Society (AGU) has requested that all academic journals under AGU should publish the original data as the paper being published. AGU announced 203 data repositories recognized by the union around the world (the website of Repository Finder), including many WDS China data centers, such as GCdataPR, WDC-RRE, Virtual Space Science Observatory, the WDC for Geophysics, and so on.

7. CoreTrustSeal certification

CoreTrustSeal (<https://www.coretrustseal.org/>) is a certification system newly released by WDS and Data Seal of Approval (DSA), based on three principal dimensionalities criteria (organization infrastructure, data management and technical capability). The CASDC and WDC-RRE in China achieved the certification at the end of 2018 and the beginning of 2019 (Wang et al. 2019). They are the first two centers getting the CoreTrustSeal certification in Asia, and presented their experiences in achieving this certification at the WDS Asia-Oceanian Conference in Beijing, 2019.

8. International Training

On August 10, 2015, the WDC-RRE hosted the international training workshop on resource and environmental data sharing in Northeast Asia and Central Asian regions in Beijing. Since then, WDC-RRE hosted training workshops for developing countries annually. More than 100 young data scientists were trained in these workshops. WDCM also has provided training courses for microbial data analysis every year. Meanwhile, WDS China supported the symposiums and conferences for Asia-Oceania regions, e.g., WDS-AO conference 2019.

9. Science popularization service

CASDC has been a member of the International Virtual Observatory Alliance (VO) since 2002. The World Wide Telescope (WWT) is a visualization environment aggregating scientific data from major telescopes, observatories, and institutions in the world. Since 2008, CASDC/China-VO has been trying to promote its application in education and science popularization in China. Hundreds of students participated and created nearly 300 great tours discussing the universe and astronomy in the WWT tour Contest. In February 2018, the CASDC/China-VO team released the first Chinese Version of WWT and the online resource sharing platform.

10. Awards

GCdataPR was awarded the 2018 WSIS Prize (champion of electronic science group) by the United Nations in March 2018 and the honorary of "leading scientific and technological achievement award -- shortlisted excellent project" by the China international big data expo in May 2018, and "innovation project" award in the eighth China digital publishing expo in July 2018. Linhuan Wu, a principal data scientist working at the

Table 1: Statistics of datasets published from June 2014 to September 2019.

Datasets Published	634	Visitors	3,057,500
Online Datasets	257GB	Data Users (IP)	45,107
Total authors (Co-authors)	935	Users from	97 countries
Author Affiliation	430	Data Files Downloaded	208,726
Authors from	12 countries	Data Downloaded	4001GB

WFCC World Data Centre for Microorganisms, won the 2017 WDS Data Stewardship Award. In addition, the CASDC received a letter of sincere gratitude from the Chinese-GEOSS for its emergency satellite data dissemination service for Iran-Iraq mag 7.8 Earthquake disaster relief in November, 2017.

4 Opportunities and Challenges of the WDS China Data Centers

Since the 1980s, China has promoted the construction of scientific data-sharing infrastructure and built scientific data centers involving many disciplines. On March 17, 2018, the State Council of China issued a circular on the Measures for Managing Scientific Data. In June 2019, the Ministry of Science and Technology and the Ministry of Finance of China implemented Measures for Managing the National Science and Technology Resource Sharing Service Infrastructure. According to the plan for the optimization and integration of the national science and technology innovation base, 20 national scientific data centers are established. These centers cover almost all of the WDS China centers and means that WDS China centers can make more progress with more funding.

Faced with the diversity of disciplines and huge data requirements from international science communities, e.g., Future Earth and Sustainable Development Goals in the UN, there are still many challenges for WDS China data centers. More work needs to be done in the near future, including metadata harvesting, data curation and archiving, repository accessibility, trustworthy certification, regional and international cooperation, training and seminars, data and popular science services, and more.

Acknowledgements

Thanks for all the materials supported by the WDS China data center community. Thanks for WDS Asia-Oceania Conference in Kyoto, 2017 and Beijing, 2019.

Funding Information

This research was funded by the National Nature Science Foundation of China (41842061), the National Science and Technology Infrastructure Center of China (2019DDJ1ZZ01), and the 13th Five-year Informatization Plan of Chinese Academy of Sciences (XXH13505-07, XXH13510-10).

Competing Interests

The authors have no competing interests to declare.

References

- Ma, JH, Duan, ZQ and Liu, C.** 2019. The “Global Change Science Research Data Publishing System” was recognized by the American Geophysical Society as the association's original journal data warehouse. *Journal of Global Change Data (Chinese and English)*, 3(3): 305–307+415–417. DOI: <https://doi.org/10.3974/geodp.2019.03.13>
- The website of Repository Finder.** URL: <https://repositoryfinder.datacite.org>.
- The website of WDS China.** URL: <http://www.wds-china.org>.
- Wang, JL, Wang, Y, Bu, K, Wang, MM and Wang, YJ.** 2019. Practice in the CoreTrustSeal Certification of World Data Center. *Journal of Agricultural Big Data*, 1(3): 69–79. DOI: <https://doi.org/10.19788/j.issn.2096-6369.190307>
- Wang, MM, Wang, JL, Zhao, Q, Gao, MS, Li, G, Han, XH and Wang, YJ.** 2017. ICPSR Scientific Data Center Construction Experience and Enlightenment. *China Science and Technology Resources Guide*, 49(6): 100–107. DOI: <https://doi.org/10.3772/j.issn.1674-1544.2017.06.015>
- Wu, LH, Sun, QL, Sugawara, H, Yang, S, Yu, GZ,** et al. 2013. Global catalogue of microorganisms (gcm): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources. *BMC Genomics*, 14: 933. DOI: <https://doi.org/10.1186/1471-2164-14-933>

How to cite this article: Wang, J, Bu, K, Wang, Y and Shao, Y. 2020. Progress in Activities of WDS-China Data Centers. *Data Science Journal*, 19: 33, pp. 1–6. DOI: <https://doi.org/10.5334/dsj-2020-033>

Submitted: 27 December 2019

Accepted: 30 July 2020

Published: 13 August 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 