

Statistics and Analysis of the Global Change Research Data Publishing & Sharing (2019)

Shi, R. X. Ma, J. H. Liu, C.* Zhang, Y. H. Shen, Y.

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

Abstract: The achievements of the Global Change Research Data Publishing & Repository (GCdataPR) are summarized in this paper concerning each of the datasets, data authors, data-related discovery papers, and data sharing. Results showed that 162 datasets covering 6 issues were published in GCdataPR in 2019, including 13,866 data files (428 compressed files). The online data size is 71.72 GB (36.41 GB after compression). Regarding the areas covered by the dataset, 5 datasets are on a global scale, 2 are trans-continental and 138 (85.19% of the total) are Asian. A total of 202 dataset authors are newly included in GCdataPR in 2019, mainly from China. Among the published datasets, 36 are directly related to data papers, and 46 research or discovery papers are directly related to datasets. There are a total of 5,205 new computer IP users, 76,510 times of downloads with a data size of 553.43 GB in 2019. The GCdataPR is playing an increasingly important role in data sharing.

Keywords: global change; data publishing; annual summary; achievement analysis; 2019

1 Introduction

In 2019, the achievements and progress on scientific data management, sharing, and policy were taken by the Chinese Academy of Sciences and national authorities in China. On February 11, 2019, the Chinese Academy of Sciences issued the “Measures for Scientific Data Management and Open Sharing of the Chinese Academy of Sciences (Trial)”^[1]. This is an important measure to implement the national big data strategy and the “Scientific Data Management Measures”. On June 5, Ministry of Science and Technology and Ministry of Finance of P. R. China jointly announced the list of 20 national scientific data centers (Guokefaji 2019 (194))^[2]. The GCdataPR has become the data publishing sub-center of the National Earth Observation Scientific Data Center. On September 26, the Chinese delegation officially released the “Earth Data Support Sustainability Report” at the 74th UN General

Received: 01-03-2020; **Accepted:** 19-03-2020; **Published:** 25-03-2020

Foundation: Chinese Academy of Sciences (XDA19090110)

***Corresponding Author:** Liu, C. L-3684-2016, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, lchuang@igsnr.ac.cn

Author ID: Shi, R. X. L-4389-2016; Ma, J. H. S-9830-2017; Liu, C. L-3684-2016; Zhang, Y. H. A-3436-2019; Shen, Y. C-6558-2019

Article Citation: Shi, R. X., Ma, J. H., Liu, C., *et al.* Statistics and analysis of the Global Change Research Data Publishing & Sharing (2019) [J]. *Journal of Global Change Data & Discovery*, 2020, 4(1): 94–100. DOI: 10.3974/geodp.2020.01.16.

Assembly^[3]. On November 1, it was put forward publicly for the first time that data can be used as production factors to participate in distribution according to their contributions, which clarified the theoretical basis for scientific data work in the new era^[4]. On November 8, the International Scientific Council Data Committee (CODATA) officially released the “The Beijing Declaration on Research Data” on its website^[5]. The Declaration affirmed the data policies and implementation progress that have been issued around the world, and on this basis, clarifies the core principles of advancing multilateral cooperation in related fields.

As of December 31, 2019, GCdataPR, as a Regular Member of the World Data System (WDS) of International Science Council (ISC) and National Earth Observation Data Publishing Center of China^[6], has published 673 datasets on 31 issues developed by 1,047 authors from 12 countries (international organizations), with a data size of 1.12 TB (258.82 GB after compression). Since August 1st, 2019, the GCdataPR has been recognized by the American Geophysical Society as one of the repositories for original data associated with academic journals of the society^[7]. To generate publicity, ensure transparency, and make the academic community clearly understand the progress of data publishing and sharing, according to items 68 and 69 of “Guidelines of Global Change Research Data Publishing & Repository”^[8], in September 2019, the Geographical Big Data Working Committee of Geographical Society of China held its 2019 annual conference in Dalian^[9]. The theme of this conference is “Geographic big data support Sustainable Development Goals.” At the conference, the Chinese Geographical Society released the “2019 global change and earth science dataset impact ranking”^[10]. This ranking was different from previous editions: the range of participating publishing systems has expanded, not limited to the GCdataPR. As the items, the impact scores rankings in four partitions for 2014–2018 datasets, data authors, and data author institutes were added besides institution ranking, foundation ranking, research paper publishing journals ranking, browsed dataset ranking, downloaded dataset ranking.

This paper summarized the data publishing result in 2019 from the perspective of datasets, dataset authors, foundations, dataset related papers, and data sharing.

2 Statistics and Analysis of Published Datasets

2.1 Published Datasets

A total of 162 datasets were published in 6 issues in 2019 (Table 1), 6 datasets less than those published in 2018^[11]. In 2019, there were 13,866 data files published in total, compressed into 428 data file packages. The compressed data file packaging rate was 32.40, and the total data size was 71.72 GB (36.41 GB after compression), with a data compressing rate of 1.97.

2.2 Geographical Regions Covered by the Datasets

Five datasets are on a global scale, accounting for 3.09% of the total (162 datasets), and two datasets are trans-continental, accounting for 1.23% of the total (Table 2). Datasets covering Asia constituted the greatest proportion (138 datasets), accounting for 85.19% of the total. Among them, 76 datasets covered China, accounting for 55.07% of the Asian datasets and 46.91% of the total published in 2019. 8 datasets covered Oceania, accounting for 4.94% of the total. The number of datasets covering Polar Regions, Europe, Africa, and North Ameri-

ca was 3, 2, 2, and 1, respectively. Also, there was 1 dataset involving philatelic culture.

Table 1 Statistics of datasets published and archived in GCdataPR in 2019

Time	Number of issues	Number of datasets	Number of data files	Number of compressed data packages	Data size (GB)	Compressed data Size (GB)
2019.01	1	20	1,261	37	0.05	0.02
2019.02	2	60	914	119	0.19	0.02
2019.03–04	3	22	1,891	40	29.24	2.29
2019.05–08	4	20	4,697	162	33.65	30.50
2019.09–11	5	20	1,314	46	7.87	3.09
2019.11–12	6	20	3,789	24	0.72	0.49
Total in 2019	6	162	13,866	428	71.72	36.41
Total during 2014–2019	31	673	411,808	1,931	1,142.34	258.82

2.3 Datasets Published by Disciplines

The datasets published in GCdataPR covered a wide range of disciplines, including geography, resources, ecology, environment, atmosphere, ocean, land, plants, water, social economy, culture, art, and history (Table 3). As shown in Table 3, there were 73 datasets about terrestrial regions (45.07%); 87 datasets about oceans, including deep-sea areas, shallow sea areas, polar regions, coastal areas, and islands (53.70%); and 2 datasets about culture and art (1.23%).

Among the terrestrial datasets (73), there were 18 datasets about water (rivers, lakes, and wetlands), accounting for 11.11% of the total, 16 datasets in humanity and economics (9.88%), 15 datasets in ecology and biology, accounting for 9.26% of the total, 11 datasets in the atmosphere (including weather and climate), accounting for 6.79%, 9 datasets about land (including land cover and land use), accounting for 5.56% of the total, and 4 in geology and geophysics (2.47%).

Table 2 Statistics of geographical regions covered by published datasets

Covering region	Number of datasets	Percentage (%)
Global	5	3.09
Trans-continental	2	1.23
Asia	138	85.19
Europe	2	1.23
North America	1	0.62
Oceania	8	4.94
Africa	2	1.23
Polar regions	3	1.85
Other (Culture)	1	0.62
Total	162	100.00

Table 3 Statistics of the domain of the published dataset

Discipline		Number of datasets	Percentage (%)
Terrestrial	Water	18	11.11
	Land	9	5.56
	Ecology/Biology	15	9.26
	Atmosphere	11	6.79
	Geology and Geophysics	4	2.47
Oceanic	Humanity/Economics	16	9.88
	Ocean (including Ocean/Coastal zone/Islands)	87	53.70
Others	Culture	2	1.23
Total		162	100.00

2.4 Data Levels

All datasets were archived into levels 0–5, each according to its stage in the developmental procedures^[12].

Based on the above criteria, 162 datasets published in 2019 were categorized into three production levels (Table 4): 87.04% of the datasets were in level 2, 9.88% of them were in level 3, and only 3.09% of the datasets were in level 4.

3 Dataset Author(s)

3.1 Dataset Author(s)

As of December 31, 2019, there were 1,047 dataset authors and 470 affiliations. Compared to the data of 2018, there are 202 new authors and 85 new affiliations.

3.2 Dataset Author Groups

Among the 162 datasets, only 4 datasets (2.47%) were developed by a single author, 141 datasets (87.04%) by a team of 2–5 persons, and 17 datasets (10.49%) by a team of more than 6 persons (Table 5).

3.3 Statistics of Chinese Authors by Affiliation and Region

3.3.1 Datasets Authors by Affiliation (Institutes or Universities)

Table 6 presents the affiliations of dataset authors. Most (72.84%) of author affiliations are from the Chinese Academy of Sciences. The second (46.3%) is from the Ministry of Natural Resources. The third (32.1%) is from the Ministry of Education. In 2019, 93 datasets (57.41%) were done by cross-departmental cooperation.

3.3.2 Chinese Authors by Region (Province, Municipality, Autonomous Region)

The distribution of dataset authors from China is shown in Table 7. Authors from Beijing constituted the highest proportion, publishing 115 datasets, accounting for 70.99% of the total datasets. The authors from Qinghai published 15 datasets. Authors from most provinces published less than 10 datasets. A total of 28 datasets were developed by trans-province authors, accounting for 17.28% of the total.

Comparing the historical data of Chinese author affiliations that published datasets^[11–12],

Table 4 Summary of the dataset in the production level

Data product level	Datasets	Percentage (%)
2	141	87.04
3	16	9.88
4	5	3.09

Table 5 Statistics of author teams and their dataset

Number of authors	Number of datasets	Percentage (%)
1	4	2.47
2–5	141	87.04
≥6	17	10.49

Table 6 Statistics of dataset author affiliations

Organization	Number of datasets	Organization	Number of datasets
Chinese Academy of Sciences	118	China Association for Science and Technology	1
Ministry of Natural Resources of P. R. China	75	Ministry of Agriculture and Rural Affairs of P. R. China	1
Ministry of Education of P. R. China	52		
China Meteorological Administration	6	Total	259
Province	3	Data published	162
Ministry of Water Resources of P. R. China	2	Dataset developed by cross-department	93
Ministry of Science and Technology of P. R. China	1	Percentage	57.41%

the authors from Beijing have the largest number of published datasets. In addition to more scientific research institutions of the global change research in Beijing, it also shows the scientists in Beijing pay more attention to data publishing and sharing. As of the end of 2019, except Hong Kong, Macao, and Taiwan, authors from 31 provinces have published datasets in China.

Table 7 Statistics of Chinese authors by region

Province	Number of datasets	Province	Number of datasets	Province	Number of datasets	Province	Number of datasets
Beijing	115	Shandong	3	Zhejiang	2	Hainan	1
Qinghai	15	Shaanxi	3	Guangxi	1	Anhui	1
Jiangsu	8	Sichuan	3	Hunan	1	Tianjin	1
Gansu	7	Shanxi	3	Ningxia	1	Fujian	1
Shanghai	5	Guizhou	2				
Guangdong	5	Henan	2		Total		198
Jilin	4	Xizang	2		Dataset published		162
Hubei	4	Jiangxi	2		Dataset developed by trans-provinces		28
Liaoning	4	Yunnan	2		Percentage		17.28%

3.4 Statistics of Datasets by Founding Agencies

Most datasets were developed with foundations (Table 8), accounting for 85.80% of the total. Notably, 14.20% of the datasets were developed by self-support, 59.26% of the datasets were funded by one funding project, and 26.54% of the datasets were funded by two or more funding projects, which usually had a large amount of data, broad coverage, and long time-series.

There were 222 funding projects (sub-projects) in the 162 published datasets. Among them, 102 projects (45.95%) were from the Chinese Academy of Sciences, 55 projects (24.77%) were from the National Natural Science Foundation of China, 29 projects (13.06%) from the Ministry of Science and Technology of P. R. China, and 22 projects (9.91%) from provincial or company support (Table 9).

Table 8 Statistics of foundation(s) supporting the dataset

Foundation	Number of datasets	Percentage (%)
No fund	23	14.20
One fund	96	59.26
More than one fund	43	26.54
Total	162	100.00

Table 9 Statistics of funding projects supporting datasets development and publishing

Foundations	Number of funding projects	Percentage (%)	Foundations	Number of funding projects	Percentage (%)
Chinese Academy of Sciences	102	45.95	National Social Science Fund of P. R. China	2	0.90
National Natural Science Foundation of China	55	24.77	China Meteorological Administration	1	0.45
Ministry of Science and Technology	29	13.06	Ministry of Civil Affairs of P. R. China	1	0.45
Province/Company	22	9.91	Outside of China	1	0.45
Ministry of Education of P. R. China	3	1.35	Others	3	1.35
Ministry of Natural Resources of P. R. China	3	1.35	Total	222	100.00

Concerning the data over the years^[11–14], the datasets supported by Chinese Academy of Sciences, National Natural Science Foundation of China, and Ministry of Science and Technology of P. R. China accounted for more than 65% of the funded datasets, which indicates that the datasets generated by national research projects are the main force for data publishing and sharing.

4 Association of Datasets, Research or Discovery Papers, and Data Papers

There are two kinds of papers associated with a dataset: a data paper and a research or discovery paper. In 2019, there were 82 research papers associated with the published datasets, including 36 data papers, 46 discovery papers, or reports. In the *Journal of Global Change Data & Discovery*, there were several columns, such as data paper, review, new data technology, data impact, data policy and strategy, EU-China cooperation, data encyclopedia, and reports on outreach. In 2019, 62 papers were published in total, including 36 data papers, 4 reviews, 1 paper on data technology, 1 on data impact scores, 2 on data policy and standards, 2 on EU-China cooperation, 11 on global change data encyclopedia, and 4 on academic activities, and 1 paper on character introduction.

5 Data Sharing Situation

From 2014 to 2019, there were 46,752 IP users from 97 countries, territories, or areas. In 2019, 5,205 IP users were newly added, and the new users were from 23 countries, territories or areas. More than 3.59 million users visited the website of the GCdataPR from 2014 to 2019 (Table 10). More than 2.25 million users visited in 2019, and the number is about five times that of 2018. From 2014 to 2019, the total number of data downloads was nearly 0.22 million by 0:00 Beijing Time (multiple downloads of the same data file within 24 hours by the same IP address were recorded as one download). The number of data downloads in 2019 is more than 70,000, which is about 4.5 times of that in 2018. From 2014 to 2019, the download data size (after compression) was more than 3.97 TB. In 2019, the download data size was 553.43 GB. The contribution of GCdataPR to the research data sharing increased year by year.

Table 10 Statistics of data sharing through the GCdataPR in 2018 and 2019*

Year	Visitors	Accum. visitors	New data users (IP)	Accum. data users (IP)	Data files downloaded	Accum. data files down-loaded	Data size downloaded (GB)	Accum. data size downloaded (GB)
2018	454,976	1,335,794	4,750	41,547	17,147	143,055	836.87	3,512.57
2019	2,256,527	3,592,321	5,205	46,752	76,510	219,565	553.43	4,066.00

*Data in 2018 is from reference [11].

6 Discussion and Conclusion

In summary, GCdataPR is steadily moving forward in 2019 and playing an increasingly important role in the publishing and sharing of scientific data. Whether it’s the number of data users, the number of dataset visits and downloads, the number of data authors, or the number of users’ countries, they are increasing year by year. Also, the number of journals of scientific papers related to the dataset is increasing. Since August 2019, the journals sponsored by

the American Geophysical Society have required authors to submit their contributions and store the dataset in the designated data center or repository. GCdataPR is honored to be one of the designated data repositories.

However, there are many issues to be discussed in the practices of data publishing. For example, some researchers are ambiguous on the concept of data intellectual property, some basic data has quality problems, some authors have low enthusiasm for contribution, etc. It is hoped that the publicity and management could be strengthened in the future, and relevant policies and mechanisms can be put forward to promote researchers to do the dataset in a down-to-earth way and make high-quality data to provide basic and important support for scientific research in global change and economic construction.

References

- [1] Scientific Data Management and Open Sharing Measures of the Chinese Academy of Sciences (Trial) [Z]. http://www.cas.cn/tz/201902/t20190220_4679797.shtml.
- [2] Notice on Issuing the List of Optimization and Adjustment of the National Science and Technology Resource Sharing Platform by the Ministry of Science and Technology and the Ministry of Finance, P. R. China [Z]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2019/201906/t20190610_147031.htm.
- [3] http://www.aircas.cas.cn/dtxw/rdxw/201909/t20190927_5402026.html.
- [4] <https://china.huanqiu.com/article/9CaKrnKnC4J>.
- [5] The Beijing declaration on research data [Z]. <https://codata.org/news/361/62/The-Beijing-Declaration-on-Research-Data>.
- [6] Liu, C., Guo, H. D., Uhler, P., *et al.* GCdataPR: infrastructure for data publishing & sharing in/for/with developing countries [J]. *Journal of Global Change Data & Discovery*, 2017, 1(1): 3–11. DOI: 10.3974/geodp.2017.01.02.
- [7] Ma, J. H., Duan, Z. Q., Liu, C. GCdataPR identified as the trusted repository by the American Geophysical Union to deposit the original data from the research paper [J]. *Journal of Global Change Data & Discovery*, 2019, 3(3): 305–307. DOI: 10.3974/geodp.2019.03.13.
- [8] Editorial Office of Journal of Global Change Data & Discovery. Guidelines of Global Change Research Data Publishing & Repository [J]. *Journal of Global Change Data & Discovery*, 2017, 1(3): 253–261. DOI: 10.3974/geodp.2017.03.01.
- [9] Zhang, W., Shen, Y. Geographic data for sciences and sustainability—summary of 2019 conference of Geographic Big Data Working Committee of Geographical Society of China [J]. *Journal of Global Change Data & Discovery*, 2019, 3(3): 308–310. DOI: 10.3974/geodp.2019.03.14.
- [10] Liu, C., Zhang, Y. H. Methodology and practice on quantifying the impact of global change & earth system science data in 2019 [J]. *Journal of Global Change Data & Discovery*, 2019, 3(3): 207–226. DOI: 10.3974/geodp.2019.03.01.
- [11] Shi, R. X., Ma, J. H., Liu, C., *et al.* Statistics and analysis of the global change research data publishing & sharing (2018) [J]. *Journal of Global Change Data & Discovery*, 2019, 3(1): 1–9. DOI: 10.3974/geodp.2019.01.01.
- [12] Shi, R. X., Liu, C., Ma, J. H., *et al.* Statistics and analysis of global change research data publishing & sharing (2014–2017) [J]. *Journal of Global Change Data & Discovery*, 2017, 1(4): 383–390. DOI: 10.3974/geodp.2017.04.01.
- [13] Geographical Society of China. Global change research data publishing & sharing rankings [R]. *Journal of Global Change Data & Discovery*, 2018, 2(3): 243–248. DOI: 10.3974/geodp.2018.03.01.
- [14] Geographical Society of China. Global change research data publishing & sharing rankings (Top 10) [J]. *Journal of Global Change Data & Discovery*, 2017, 1(2): 249–251. DOI: 10.3974/geodp.2017.02.23.