

全球变化科学研究数据 2019 年度出版成果分析

石瑞香, 马军花, 刘 闯*, 张应华, 申 艳

中国科学院地理科学与资源研究所, 北京 100101

摘 要: 本文从出版的数据集、数据作者、数据关联论文和数据共享四方面, 对 2019 年出版成果做以总结与分析。统计表明, 2019 年“全球变化科学研究数据出版系统”总计出版了 6 期 162 个数据集, 出版数据文件 13,866 个, 压缩为 428 个数据文件包, 出版数据量 71.72 GB, 压缩后出版量为 36.41 GB。数据集覆盖区域上, 亚洲的数据集出版最多(138 个), 占年度总出版数据集的 85.19%; 覆盖全球的数据集 5 个; 跨洲区域的数据集 2 个。2019 年新增 365 篇数据论文、46 篇科学发现论文或研究报告与数据集相关联。新增数据集作者 202 位, 新增作者单位 85 个。新增 5,205 个 IP 用户, 数据下载次数增加 76,510 次、下载量增加 553.43 GB。“全球变化科学研究数据出版系统”在数据共享方面正在发挥越来越重要的作用。

关键词: 全球变化; 科学研究数据; 数据出版; 2019

DOI: 10.3974/geodp.2020.01.16

1 背景

2019 年是“全球变化科学研究数据出版系统”(Global Change Research Data Publishing & Repository)正式开通以来的第六年。这一年, 国家部委、中国科学院等在科学数据管理、共享和政策等方面都有了阶段性进展。2019 年 2 月 11 日, 中国科学院印发《中国科学院科学数据管理与开放共享办法(试行)》^[1]。这是落实国家大数据战略和《科学数据管理办法》的重要举措。6 月 5 日, 科技部、财政部联合公布了 20 个国家科学数据中心名单(国科发基 2019(194)号^[2])。“全球变化科学研究数据出版系统”成为国家对地观测科学数据中心的数据出版分中心。9 月 26 日, 在 74 届联合国大会上, 中国代表团正式发布《地球大数据支撑可持续发展报告》^[3]。在 11 月 1 日召开的中共十九届四中全会上, 首次公开提出数据可作为生产要素按贡献参与分配, 为新时期科学数据工作明确了理论基础^[4]。11 月 8 日, 国际科学理事会数据委员会(CODATA)在其官方网站正式发布《科研数据北京宣言》^[5]。《宣言》肯定了世界各地已发布的数据政策和实施进展, 并在此基础上阐明了推进相关领域多边合作的核心原则。

截至 2019 年 12 月 31 日, “全球变化科学研究数据出版系统”作为发展中国家数据出版基础设施^[6], 共计出版了 31 期 673 个数据集, 出版数据量 1.12 TB(压缩后 258.82 GB)。这些数据来自 12 个国家(或国际组织)的 1,047 位作者。自 2019 年 8 月 1 日起, “全球

收稿日期: 2020-03-01; 修订日期: 2020-03-19; 出版日期: 2020-03-25

基金项目: 中国科学院(XDA19090110)

*通讯作者: 刘闯 L-3684-2016, 中国科学院地理科学与资源研究所, lchuang@igsnr.ac.cn

作者 ID: 石瑞香 L-4389-2016, 马军花 S-9830-2017, 刘闯 L-3684-2016, 张应华 A-3436-2019, 申艳 C-6558-2019

引用格式: 石瑞香, 马军花, 刘闯等. 全球变化科学研究数据 2019 年度出版成果分析[J]. 全球变化数据学报, 2020, 4(1): 94–100. DOI: 10.3974/geodp.2020.01.16.

变化科学研究数据出版系统”被美国地球物理学会认定为该学会学术期刊关联原创数据的仓储中心^[7]。根据《全球变化科学研究数据出版与共享指南》第六十八条和第六十九条规定^[8]，2019 年 9 月，中国地理学会地理大数据工作委员会 2019 年会在大连召开^[9]。这次会议的主题是“地理大数据支持可持续发展目标”。会上，中国地理学会发布了“2019 年全球变化暨地球科学数据集影响力排行榜”^[10]。这次排行榜与往届不同的是：参与的出版系统范围扩大，不只限于“全球变化科学研究数据出版系统”；在发布的排行榜内容上，除了数据集作者单位排行榜、基金资助排行榜、学术期刊论文关联原创实体数据出版期刊排行榜、热点数据集网络浏览量排行榜、热点数据集网络下载次数排行榜，还增加了 2014–2018 年数据集、数据作者、数据作者单位的四个分区的影响力积分排行榜。

本文从出版的数据集、数据集作者、数据集资助基金、数据集关联论文和数据共享等方面对 2019 年度数据出版工作进行统计和总结。

2 数据集成果统计

2.1 数据集成果

2019 年“全球变化科学研究数据出版系统”总计出版了 6 期 162 个数据集（表 1），出版数量与 2018 年出版量基本持平^[11]。2019 年，出版数据文件总数为 13,866 个，压缩为 428 个数据文件包，文件打包率为 32.40。出版数据总量为 71.72 GB，压缩为 36.41 GB，数据压缩比为 1.97。

表 1 2019 年“全球变化科学研究数据出版系统”出版数据集统计表

时间	期号	出版数据集数（个）	出版数据文件数（个）	数据文件包（个）	出版数据量（GB）	压缩后数据量（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
2019 合计	6	162	13,866	428	71.72	36.41
2014–2019	31	673	411,808	1,931	1,142.34	2,58.82

2.2 数据集覆盖的地理区域

2019 年“全球变化科学研究数据出版系统”出版的数据集中，有 5 个数据集覆盖全球，占本年度出版数据集总数的 3.09%（表 2）。跨洲区域数据集 2 个，占本年度出版数据集总数的 1.23%。覆盖区域位于亚洲的数据集出版最多，达 138 个，占本年度出版数据集的 85.19%。其中，中国区域的数据集为 76 个，占亚洲区域数据集的 55.07%，占 2019 年度出版数据集的 46.91%。大洋洲数据集 8 个，占本年度出版数据集的 4.94%。有关极地的数据集有 3 个，有关欧洲、非洲的数据集分别有 2 个，有关北美洲的数据集有 1 个。此外，还有 1 个涉及集邮文化的数据集。

2.3 数据集学科领域

全球变化科学研究数据涉及的学科领域比较广泛，包括：地理、资源、生态、环境、大气、海洋、土地、植被、水域、社会经济、人文、艺术等。表 3 列出了 2019 年“全球变化科学研究数据出版系统”出版的数据集所在领域及其比例。在 2019 年出版的数据集中，学科领域涉及陆地的数据集 73 个，占本年度出版数据集的 45.07%；涉及海洋（包括深海、浅海、极地、海岸带和海岛等）的数据集 87 个，占本年度出版数据集的 53.70%。此外，文化、艺术领域 2 个，占本年度出版数据集的 1.23%。

在出版的陆地数据集中，水文（包括河流、湖泊、湿地等）数据集最多，为 18 个，占本年度出版数据集的 11.11%；其次为人文、经济领域，出版了 16 个数据集，占本年度出版数据集的 9.88%。再次是生态、生物领域数据集 15 个，占比 9.26%。大气领域（包括天气、气候等）数据集 11 个，占本年度出版数据集的 6.79%。

土地（包括土地覆盖，土地利用等）数据集 9 个，占 2019 年出版数据集的 5.56%；地质和地球物理领域出版数据集 4 个，占本年度出版数据集的 2.47%。

2.4 数据集级别

根据国内外地球观测数据及全球变化科学研究数据从获取到数据产品研发这一过程，将数据集划分为 0-5 级，数据集级别划分的基本原则和标准见文献^[12]。

对 2019 年“全球变化科学研究数据出版系统”出版的 162 个数据集分别予以分级并汇总。由表 4 知，大部分数据集为基础性 2 级数据，占总出版数据集的 87.04%；3 级数据集 16 个，占比 9.88%；4 级数据集 5 个，占比 3.09%。

表 2 数据集覆盖地理区域统计表

覆盖区域	数据集个数	占百分比（%）
全球	5	3.09
跨洲区域	2	1.23
亚洲	138	85.19
欧洲	2	1.23
北美洲	1	0.62
大洋洲	8	4.94
非洲	2	1.23
极地	3	1.85
其他（文化）	1	0.62
总计	162	100.00

表 3 数据集涉及学科领域统计表

学科领域		数据集个数	占百分比（%）
陆 地	水文	18	11.11
	土地	9	5.56
	生态、生物	15	9.26
	大气	11	6.79
	地质与地球物理	4	2.47
海 洋	人文、经济	16	9.88
	海洋（包括海洋、海岸带、岛屿）	87	53.70
其他	文化、艺术等	2	1.23
总计		162	100.00

表 4 数据集分级汇总表

所属分级	出版数据集	占百分比（%）
2	141	87.04
3	16	9.88
4	5	3.09

3 数据集作者与单位

3.1 数据集作者

截至 2019 年 12 月 31 日,“全球变化科学研究数据出版系统”数据集作者达 1,047 位,作者单位达 470 个。与 2018 年相比,新增数据集作者 202 位,新增作者单位 85 个。

3.2 数据集作者人数

在 2019 年出版的数据集中,有 4 个数据集由独立作者完成,占本年度出版数据集的 2.47%;大部分数据集(141 个)是由 2-5 人组成的团队完成,占本年度出版数据集的 87.04%;由 6 人以上组成的团队研发、出版的数据集 17 个,占本年度出版数据集的 10.49%(表 5)。

3.3 中国作者单位所在部门和地区

3.3.1 中国作者单位所在部门

表 6 列出了 2019 年出版数据的中国作者所在单位隶属的部门。其中,中国科学院的数据作者最多,独立或参与出版 118 个数据集,占当年出版数据集的 72.84%;其次是自然资源部的作者,独立或参与出版了 75 个数据集,占当年出版数据集的 46.30%;再次是教育部系统的作者,独立或参与出版了 52 个数据集,占当年出版数据集的 32.10%。与 2018 年(跨系统合作数据集占当年出版数据集的 27.98%)相比^[12],跨部门合作完成的数据集明显增多,达到 93 个,占 2019 年出版数据集的 57.41%。

表 5 按作者人数划分的数据集统计表

作者人数(人)	数据集个数	占百分比(%)
1	4	2.47
2-5	141	87.04
≥6	17	10.49

表 6 按中国作者所在部门划分的数据集统计表

作者单位 所在部门	数据集 个数	作者单位 所在部门	数据集 个数
中国科学院	118	中国科学技术协会	1
自然资源部	75	农业农村部	1
教育部	52		
中国气象局	6	总计	259
地方	3	实际出版数据集	162
水利部	2	跨系统合作数据集	93
科学技术部	1	跨系统合作数据集比例	57.41%

3.3.2 中国作者所在地(省、直辖市、自治区)

2019 年出版数据集的中国作者单位所在省(直辖市、自治区)的情况列于表 7。来自北京的作者出版数据集最多,达 115 个,占本年度出版数据集的 70.99%;其次是来自青海的作者,出版了 15 个数据集;多数省份的作者出版的数据集在 1-10 之间。在出版的数据集中,有 28 个数据集是跨省(直辖市、自治区)合作完成的,占本年度出版数据集的 17.28%。

比较出版数据集的中国作者单位的历年数据可知^[11-12],来自北京的作者出版数据集数量最多,除了全球变化科学数据相关领域在京科研单位较多这一原因外,也说明北京的科研单位对科学数据出版的关注度高,对数据共享的认知度高,对数据出版这一科学事业的参与度高。截至 2019 年末,除港澳台以外,中国的 31 个省(直辖市、自治区)均有作者出版了数据集。

表 7 按中国作者所在地划分的数据集统计表

作者 所在地	数据集 个数	作者 单位	数据集 个数	作者 单位	数据集 个数	作者 单位	数据集 个数
北京	115	山东	3	浙江	2	海南	1
青海	15	陕西	3	广西	1	安徽	1
江苏	8	四川	3	湖南	1	天津	1
甘肃	7	山西	3	宁夏	1	福建	1
上海	5	贵州	2				
广东	5	河南	2			合计	198
吉林	4	西藏	2			实际出版数据集	162
湖北	4	江西	2			跨地区合作出版数据集	28
辽宁	4	云南	2			跨地区合作出版数据集比例	17.28%

3.4 数据集的资助基金

对出版数据集的资助基金进行统计得到，大部分数据集研发都有基金项目资助（占 85.80%）。有 14.20%的数据集是学者自由选题研发的。在有基金资助的数据集中，由 1 个基金资助的数据集占当年总出版数据集的 59.26%，由 2 个及以上基金资助研发的数据集占当年总出版数据集 26.54%（表 8）。这部分多为数据量大、覆盖区域广、数据集级别较高，或时间序列较长的数据集。

表 8 按资助基金项数划分的数据集统计表

资助基金	数据集个数	所占比例（%）
没有基金资助	23	14.20
1个基金资助	96	59.26
2个及以上基金资助	43	26.54
总计	162	100.00

在出版的 162 个数据集中，标注有 222 项基金项目（课题）资助研发和出版。在这些基金项目（课题）中，有 102 项来自中国科学院，占基金项目总数的 45.95%；55 项来自国家自然科学基金，占基金项目总数的 24.77%；有 29 项来自国家科学技术部，占基金项目总数的 13.06%；有 22 项来自地方或企业，占基金项目总数的 9.91%（表 9）。

表 9 资助数据集研发与出版的各类基金项目（课题）统计表

基金项目来源	基金项目 (课题)数	占 百 分 比 (%)	基金项目来源	基金项目 (课题)数	占百分比 (%)
中国科学院	102	45.95	国家社会科学基金	2	0.90
国家自然科学基金	55	24.77	中国气象局	1	0.45
科技部	29	13.06	民政部	1	0.45
地方、企业	22	9.91	国外	1	0.45
教育部	3	1.35	其它	3	1.35
自然资源部	3	1.35	总计	222	100.00

比较出版数据集资助基金的历年数据可知^[11-14]，中国科学院、国家自然科学基金、国家科学技术部资助项目出版的数据集占到有基金资助数据集的 65% 以上，表明在数据集出版方面，国家级科研项目资助产生的数据集是数据共享的主力军。

4 实体数据与数据信息关联

与实体数据关联的论文包括两部分：其一是与实体数据关联的数据论文；其二是作者发表的与实体数据直接关联的科学发现论文。2019 年实体数据新增关联 82 篇论文，包括 36 篇数据论文、46 篇科学发现论文或研究报告。2019 年，在《全球变化数据学报》（中英文）上，开设有数据论文、综述、新技术、数据科学影响力、数据政策与科学计划、中欧合作、全球变化数据大百科辞条和学术活动等栏目，共出版 62 篇论文，其中数据论文 36 篇，综述 4 篇，新技术 1 篇，数据科学影响力 1 篇，数据政策与标准规范 2 篇，中欧合作 2 篇，全球变化数据大百科辞条 11 个，学术活动报道 4 篇，人物介绍 1 篇。

5 数据共享统计

2014–2019 年，共有 46,752 个计算机 IP 用户，这些用户来自 97 个国家或地区。2019 年新增 5205 个 IP 用户，用户范围新增 23 个国家或地区。2014–2019 年，“全球变化科学研究数据出版系统”网站访问累计达 359 万余人次（表 10）。2019 年新增访问 225 万余人次，2019 年网站访问人次约是 2018 年的 5 倍。2014–2019 年，数据下载次数累计近 22 万（以北京时间零时为基准，24 小时内同一台计算机 IP 用户多次下载同一个数据文件，按一次记录）。2019 年数据下载 7 万余次，约是 2018 年数据下载次数数的 4.5 倍。2014–2019 年，累计数据下载量（压缩后）在 3.97 TB 以上。2019 年，新增数据下载量 553.43 GB。可见，“全球变化科学研究数据出版系统”在科学数据共享方面的影响在逐渐增加。

表 10 2018、2019 年全球变化科学研究数据共享统计表*

年	访问 人次	累计访问 人次	新增用户数 (IP)	累计用 户数 (IP)	数据下载 次数	累计下载 次数	数据下载量 (GB)	累计下载量 (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

* 2018 年数据来自参考文献[11]。

6 讨论与结论

由上面的统计和分析知，2019 年“全球变化科学研究数据出版系统”正在稳步前行，在科学数据的出版与共享方面起着越来越重要的作用。无论是数据用户数、数据集的访问量、下载量，还是数据作者的人数、用户所在国家数等等，都在逐年增加。另外，数据关联的科学论文所在期刊也在逐渐增多。2019 年 8 月以来，美国地理物理学会主办的期刊开始要求作者投稿的同时，将数据集存储在指定的数据仓储中心。“全球变化科学研究数据出版系统”很荣幸成为指定的数据仓储之一。

在“全球变化科学研究数据出版系统”取得成绩的同时,在数据出版实践中,也遇到许多问题有待商讨和解决,如数据知识产权概念模糊、部分基础数据的质量控制、数据作者投稿的积极性等。有关数据共享的价值与意义、现状与未来还需要加强管理和宣传,并呼吁将数据成果纳入科研工作者的考评体系和绩效体系,出台有关的政策和有效的激励机制,促进科研人员踏踏实实做数据,做高质量的数据,为全球变化科学研究和经济建设提供基础和支撑。

参考文献

- [1] 中国科学院科学数据管理与开放共享办法(试行)[Z]. http://www.cas.cn/tz/201902/t20190220_4679797.shtml.
- [2] 科技部 财政部关于发布国家科技资源共享服务平台优化调整名单的通知[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] 科研数据北京宣言[Z]. <https://codata.org/news/361/62/The-Beijing-Declaration-on-Research-Data>.
- [6] 刘闯,郭华东,Uhlir, P. F.等.发展中国家数据出版基础设施与共享政策研究[J].全球变化数据学报,2017,1(1): 3-11. DOI: 10.3974/geodp.2017.01.02.
- [7] 马军花,段宗奇,刘闯.“全球变化科学研究数据出版系统”被美国地球物理学会认定为该学会学术期刊关联原创数据仓储中心[R].全球变化数据学报,2019,3(3): 305-307. DOI: 10.3974/geodp.2019.03.13.
- [8] 《全球变化数据学报》(中英文)编辑部.全球变化科学研究数据出版与共享指南[J].全球变化数据学报,2017,1(3): 253-261. DOI: 10.3974/geodp.2017.03.01.
- [9] 张威,申艳.地理大数据支持可持续发展目标大会——中国地理学会地理大数据工作委员会 2019 年会[R].全球变化数据学报,2019,3(3): 308-310. DOI: 10.3974/geodp.2019.03.14.
- [10] 刘闯,张应华.全球变化暨地球科学数据影响力分区方法及 2019 年实践[J].全球变化数据学报,2019,3(3): 207-226. DOI: 10.3974/geodp.2019.03.01.
- [11] 石瑞香,马军花,刘闯等.全球变化科学研究数据出版与共享成果分析(2018) [J].全球变化数据学报,2019,3(1): 1-9. DOI: 10.3974/geodp.2019.01.01.
- [12] 石瑞香,刘闯,马军花等.全球变化科学研究数据出版成果分析(2014-2017) [J].全球变化数据学报,2017,1(4): 383-390. DOI: 10.3974/geodp.2017.04.01.
- [13] 中国地理学会.全球变化科学研究数据出版与共享排行榜[R].全球变化数据学报,2018,2(3): 243-248. DOI: 10.3974/geodp.2018.03.01.
- [14] 中国地理学会.全球变化科学研究数据出版与共享排行榜(前 10 名) [R].全球变化数据学报,2017,1(2): 249-251. DOI: 10.3974/geodp.2017.02.23.