

全球变化科学研究数据出版成果分析 (2014–2017)

石瑞香*, 刘 闯, 马军花, 王正兴

中国科学院地理科学与资源研究所, 北京 100101

摘 要:“全球变化科学研究数据出版系统”是国际科学理事会(ICSU)世界数据系统(WDS)首个实体数据与数据论文关联出版的系统,是中国地球观测组织(China GEO)数据出版分中心,是科睿唯安数据被检索出版系统(DCI)。本文从出版的数据集、数据作者、数据关联论文、数据共享四方面,对2014–2017年出版成果做以总结与分析,并对工作中存在的一些问题进行了讨论。成果统计表明,2014–2017年,“全球变化科学研究数据出版系统”总计出版了17期343个数据集。出版数据文件394,026个,压缩为1,226个数据文件包;出版数据量1,028.13 GB,压缩为209.10 GB。数据集覆盖区域上,亚洲的数据集最多(248个),占总出版数据集的72.30%;覆盖全球的数据集11个;跨洲区域(包括一带一路)的数据集23个。数据集作者总计567位,他们来自11个国家,包括:中国、日本、美国、泰国等。其中,中国作者占比达94.89%。在出版的数据集中,有93个数据集与数据论文直接关联,有119篇科学发现论文与数据集直接关联。来自76个国家、3.68万台计算机IP用户浏览和下载了出版的数据。网页浏览量达88.08万人次,数据下载12.59万次,下载量2,675.70 GB。其中,中国是主要用户(32,749),占全部用户的89%,其次是美国、澳大利亚、日本、马来西亚等国。“全球变化科学研究数据出版系统”获得中国大数据优秀产品、服务和应用解决方案最佳案例(2016)。

关键词: 全球变化; 科学研究数据; 数据出版; 2014–2017

DOI: 10.3974/geodp.2017.04.01

1 背景

自全球变化科学研究计划^[1–3]实施以来,与全球变化科学研究相关的科学数据不断产生。为了推动全球变化科学研究领域科学数据成果的出版、保藏、传播和共享,中国科学院地理科学与资源研究所、中国地理学会于2014年6月正式开通“全球变化科学研究数据出版系统”(Global Change Research Data Publishing & Repository)。截至2017年12月20日,共计出版了17期343个数据集。为了确保这些出版的数据公开、透明,并使学术界全面了解数据出版和共享的进展情况,《全球变化数据学报》(中英文)编辑部根据“全球变化科学研究数据出版与共享指南”^[4]的基本要求,对数据出版的进展进行年度总结并向主办单位、资助单位、合作伙伴单位等做出报告。总结分两种方式进行:其一是发布排行榜,其二是综合性总结。“全球变化科学研究数据出版与共享2017年排行榜”已经于2017年6

收稿日期: 2017-12-23; 修订日期: 2017-12-25; 出版日期: 2017-12-25

基金项目: 中国科学院地理科学与资源研究所(2017); 中国地理学会(2017); 中国科学院(2017)

*通讯作者: 石瑞香 L-4389-2016, 中国科学院地理科学与资源研究所, shirx@igsrr.ac.cn

作者ID: 石瑞香 L-4389-2016, 刘闯 L-3684-2016, 马军花 S-9830-2017, 王正兴 L-5255-2016

引用格式: 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.

月在中国首次全球变化科学研究数据出版与共享大会上发布^[5-6]。其中包括数据集作者单位、数据出版资助基金项数、学术期刊论文关联原创实体数据出版期刊、热点数据集网络浏览次数和热点数据集网络下载次数的排行榜。本文是全球变化科学研究数据出版与共享综合性总结，涉及的时间是 2014 年 6 月 1 日-2017 年 12 月 20 日。

2 数据集成果统计

2.1 数据集成果

2014-2017 年，“全球变化科学研究数据出版系统”总计出版了 17 期 343 个数据集（表 1）。其中，2014 年出版了两期共 36 个数据集；2015 年出版了两期共 34 个数据集；2016 年出版了 9 期共 190 个数据集；2017 年配合《全球变化数据学报》（中英文）的创刊，出版了 4 期共 83 个数据集。出版数据文件总量为 394,026 个，压缩为 1,226 个数据文件包，文件打包率为 321.39。出版数据总量为 1,028.13 GB，压缩为 209.10 GB，数据压缩率为 4.92。

表 1 2014-2017 年“全球变化科学研究数据出版系统”出版数据集统计表

年/月	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)
2014.06	1	20	17,843	57	4.58	3.95
2014.12	2	16	15,961	99	71.54	18.90
小计	2	36	33,804	156	76.12	22.85
2015.04-06	1	18	49,771	166	103.19	35.00
2015.07-12	2	16	4,964	130	35.78	20.80
小计	2	34	54,735	296	138.97	55.80
2016.01-03	1	18	2,090	35	3.71	1.36
2016.04-05	2	20	409	26	0.86	0.26
2016.06	3	21	871	83	48.78	9.95
2016.07	4	21	406	42	19.94	1.00
2016.08	5	21	470	74	8.80	7.24
2016.09	6	25	1,499	46	0.11	0.01
2016.10	7	25	595	51	0.65	0.24
2016.11	8	21	251	35	1.11	0.33
2016.12	9	18	402	35	0.34	0.16
小计	9	190	6,993	427	84.30	20.55
2017.01-02	1	21	355	93	25.67	14.80
2017.03-06	2	20	1,100	63	39.37	5.08
2017.07-10	3	20	7,662	32	4.98	0.32
2017.11-12	4	22	289,377	159	658.72	89.70
小计	4	83	298,494	347	728.74	109.90
合计	17	343	394,026	1,226	1,028.13	209.10

2.2 数据集覆盖的地理区域

“全球变化科学研究数据出版系统”出版的数据集中，有 11 个数据集覆盖全球，占总

出版数据集的 3.21%（表 2）。跨洲区域（包括一带一路）数据集 23 个，占总出版数据集的 6.71%。亚洲的数据集出版最多，达 248 个，占总出版数据集的 72.30%。其中，中国区域的数据集为 148 个，占亚洲区域数据集的 59.68%，占总出版数据集的 43.15%。欧洲的数据集 21 个，占总出版数据集的 6.12%；北美洲的数据集 16 个，占总出版数据集的 4.66%；拉丁美洲、大洋洲和非洲的数据集都很少。南北极地区的数据集 11 个，占总出版数据集的 3.21%。此外，还有 3 个涉及数据方法论和新技术的数据集。

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、历史等。表 3 列出了“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。在全部数据集中，学科领域涉及陆地的数据集 228 个，占总出版数据集的 66.47%；海洋（包括深海、浅海、极地、海岸带和海岛等）数据集 108 个，占总出版数据集的 31.49%。此外，文化艺术领域（例如集邮艺术品中的南森博士数据集等）4 个，占总出版数据集的 1.17%；技术、方法领域的数据集 3 个，占 0.87%。

在出版的陆地数据集中，水域（包括河流、湖泊、湿地等）数据集 40 个，占总出版数据集的 11.66%；土地（包括土地覆盖，土地利用等）数据集 24 个，占总出版数据集的 7.00%。在生态、生物领域出版数据集最多，为 79 个，占 23.03%。大气领域（包括天气、气候等）数据集 40 个，占总出版数据集的 11.66%。地质矿产、环境领域均出版数据集 7 个，分别占总出版数据集的 2.04%；灾害数据集 5 个，占总出版数据集的 1.46%。人文方面的数据出版了 26 个，占总出版数据集的 7.58%。

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获取至数据产品研发过程，数据集划分 0–5 级，数据集级别划分的基本原则为：

0 级：机器获得的数据或信号

表 2 数据集覆盖地理区域统计表

研究区域	数据集个数	占百分比 (%)
全球	11	3.21
跨洲区域（包括一带一路）	23	6.71
亚洲	248	72.30
欧洲	21	6.12
北美洲	16	4.66
拉丁美洲	4	1.17
大洋洲	3	0.87
非洲	3	0.87
极地	11	3.21
技术、方法	3	0.87
总计	343	100

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	水域	40	11.66
	土地	24	7.00
	生态、生物	79	23.03
	大气	40	11.66
	地质矿产	7	2.04
	环境	7	2.04
	灾害	5	1.46
	人文	26	7.58
海 洋	海洋（包括海洋、 海岸带、岛屿）	108	31.49
其 他	文化艺术	4	1.17
	技术、方法	3	0.87
总 计		343	100

- 1 级：在 0 级数据基础上，经过几何校正、辐射校正等处理的数据
- 2 级：1 级数据+智力投入后产生的数据集
- 3 级：2 级数据+智力投入后产生的数据集
- 4 级：多个 2 级或 3 级数据+智力投入后产生的数据集
- 5 级：多个 2 级或 3 级或 4 级数据+智力投入后产生的综合性、时间序列、覆盖全球数据集

根据这一划分原则，对“全球变化科学研究数据出版系统”出版的 343 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 72.89%；3 级数据集 84 个，占 24.49%；4 级数据集 9 个，占 2.62%。

3 数据集作者与单位

3.1 数据集作者

数据集作者（包括第一作者、通讯作者、合作作者）共计 567 位，他们来自 11 个国家：中国、日本、美国、泰国、俄罗斯、智利、巴基斯坦、肯尼亚、马达加斯加、荷兰和捷克。其中，中国作者 538 位，占 94.89%。

由表 5 知，中国作者出版的数据集有 337 个，占总出版数据集的 98.25%；其次为日本作者，出版了 13 个数据集。国内外作者合作出版数据集 16 个，占总出版数据集的 4.66%。国外作者独立发表的数据集有 6 个，占总出版数据集的 1.75%。

3.2 数据集作者人数

在出版的数据集中，有 25 个数据集由独立作者完成，占总出版数据集的 7.29%；大量的数据集（257 个）是由 2-5 人组成的小团队完成，占总出版数据集的 74.93%；

由 6-10 人组成的团队研发、出版的数据集 56 个，占总出版数据集的 16.33%。有 5 个数据集是由 10 位以上作者合作完成（表 6）。作者最多的数据集合作作者数量为 33 人。

3.3 中国作者所在部门和地区

3.3.1 中国作者所在部门

表 7 列出了中国作者所在单位隶属的部门。中国科学院的数据作者最多，其次是教育部系统。虽然，数据集作者以本部门合作为主，跨部门合作完成的数据集也达到 79 个，占 23.03%，这种跨部门数据集研发和出版的合作是乐观的。

表 4 数据集分级汇总表

出版数据集	所属分级	占百分比 (%)
250	2	72.89
84	3	24.49
9	4	2.62

表 5 按作者国别划分的数据集统计表

作者国别	数据集个数	作者国别	数据集个数
中国	337	荷兰	1
日本	13	捷克	1
美国	5		
泰国	2	总计	364
俄罗斯	1	实际出版数据集	343
智利	1	国外作者独立发表数据集	6
巴基斯坦	1	国外作者独立发表数据集比例	1.75%
肯尼亚	1	跨国合作数据集	16
马达加斯加	1	跨国合作数据集比例	4.66%

表 6 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	25	7.29
2-5	257	74.93
6-10	56	16.33
>10	5	1.46

3.3.2 中国作者所在地（省、直辖市、自治区）

据统计,中国作者来自 138 个单位,作者单位分布各省(直辖市、自治区)的情况列于表 8。来自北京的作者出版数据集最多,达 256 个;上海、四川、河南、甘肃和江苏五省、市作者出版的数据集在 10-20 之间;多数省份出版的数据集为个位数;宁夏、青海、贵州、香港、澳门、台湾一些省份和地区没有作者出版数据。在出版的数据集中,以本省区合作为主,占全部数据集的 83.80%,而跨省(市、自治区)合作的数据集仅占 17.20%。

表 7 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	240	中国地震局	2
教育部	128	农业部	2
国家测绘地理信息局	9	地方	2
国家海洋局	8	水利部	1
中国气象局	8	国家发展与改革委员会	1
企业	8	总计	422
国家林业局	7	实际出版数据集	343
国土资源部	4	跨系统合作数据集	79
住房和城乡建设部	2	跨系统合作数据集比例	23.03%

表 8 按中国作者所在地划分的数据集统计表

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	作者单位	数据集个数
北京	256	陕西	8	湖北	3	海南	1
上海	16	广东	7	江西	3	安徽	1
四川	15	山东	6	福建	3	天津	1
河南	13	新疆	5	湖南	3	西藏	1
甘肃	12	山西	5	云南	2	重庆	1
江苏	12	河北	5			合计	402
吉林	8	内蒙古	4			实际出版数据集	343
辽宁	4	黑龙江	2			跨地区合作出版数据集	59
浙江	3	广西	2			跨地区合作出版数据集比例	17.20%

3.4 数据集的资助基金

由出版数据集的资助基金统计知,大部分数据集研发都有基金项目资助(占 84.55%)。但是,仍有 15.45%的数据集是学者自由选题研发的。在由基金资助的数据集中,由一个基金资助的数据集占总出版数据集的 47.23%,由 2 个及 2 个以上基金资助研发的数据集占 37.32%(表 9)。这部分多为数据量很大,覆盖区域很广,数据集级别较高,或时间序列较长的数据集。

表 9 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	53	15.45
1个基金资助	162	47.23
2个及以上基金资助	128	37.32
总计	343	100

在出版的 343 个数据集中,标注有 532 项基金项目(课题)资助研发、出版。其中,有 139 项来自国家科学技术部,占基金项目总数的 26.13%;有 123 项来自中国科学院,占基金项目总数的 23.12%;有 118 项来自国家自然科学基金,占项目总数的 22.18%(表 10)。

4 实体数据与数据信息关联

为了确保科学研究数据具有信息全面、可再用，实体数据与足够详实的数据信息关联出版非常必要^[7]。为此，中国科学院地理科学与资源研究所和中国地理学会于 2017 年 3 月创办了《全球变化数据学报》(中英文)。该学报主要刊载与“全球变化科学研究数据出版系统”出版的实体数据关联的数据论文，也发表关于学科发展的综述、政策、技术等方面的论文和学术活动报道。此外，主办单位于 2016 年启动了“全球变化科学研究数据出版系统”学术期刊合作伙伴计划。截至 2017 年底，37 种学术期刊加入该团队。主要目的是为各个期刊发表学术论文的同时，提供论文关联数据的出版平台；从而在关联数据方面为进一步提高合作伙伴学术期刊质量和影响力开辟一条新的途径。“全球变化科学研究数据出版系统”是国际科学理事会(ICSU)世界数据系统(WDS)首个实体数据与数据论文关联出版的系统，是中国地球观测组织(China GEO)数据出版分中心，是科睿唯安数据被检索出版系统(DCI)^[8]。被评为中国大数据优秀产品、服务和应用解决方案最佳案例(2016)^[9]。

表 10 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(%)
科技部	139	26.13	发改委	2	0.38
中国科学院	123	23.12	环保部	2	0.38
国家自然科学基金	118	22.18	国家林业局	2	0.38
地方、企业	91	17.11	中国国家航天局	2	0.38
教育部	19	3.57	中国地震局	2	0.38
国家海洋局	10	1.88	水利部	1	0.19
国家社会科学基金	6	1.13	人事部	1	0.19
国土资源部	3	0.56	国家旅游局	1	0.19
中国气象局	2	0.38			
国外	8	1.50	总计	532	100

数据关联论文包括两方面：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。2014 年数据论文是以《地理学报》增刊的方式发表，直至《全球变化数据学报》(中英文)创刊，数据论文才陆续增加。由表 11 知，近 4 年来，共出版 93 篇数据论文，实体数据关联 119 篇科学发现论文。数据出版的基础设施建设逐步完善，2016 年及以前出版的实体数据关联的数据论文滞后发表并逐渐补齐。

5 数据共享统计

5.1 数据共享量

表 12 列出了全球变化科学研究数据历年和累计共享量。其中，网站访问累计达 88 万余人次，计算机 IP 用户达 3 万余，数据下载次数 12 万余(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)，数据下载量(压缩后)在 2TB 以上。2014 年以来，计算机 IP 用户数、数据下载次数每年都在稳步增多(图 1，图 2)。

表 11 数据论文与科学发现论文统计表

年份	数据集 个数	数据论文 篇数	科学发现 论文篇数	论文总计
2014	36	20	1	21
2015	34	0	0	0
2016	190	0	95	95
2017	83	73	23	96
总计	343	93	119	212

5.2 数据用户的地理分布

数据用户来自亚洲、欧洲、大洋洲、北美洲、南美洲、非洲的 76 个国家。主要用户是中国用户（32,749），占全部用户的 89%。其次是美国、澳大利亚、日本、马来西亚等国（图 3）。

表 12 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)
2014	332,846	332,846	174	174	822	822	25.79	25.79
2015	124,668	457,514	9,764	9,938	23,726	24,548	976.11	1,001.90
2016	339,870	797,384	10,701	20,639	47,867	72,415	703.31	1,705.21
2017	83,434	880,818	16,158	36,797	53,493	125,908	970.49	2,675.70

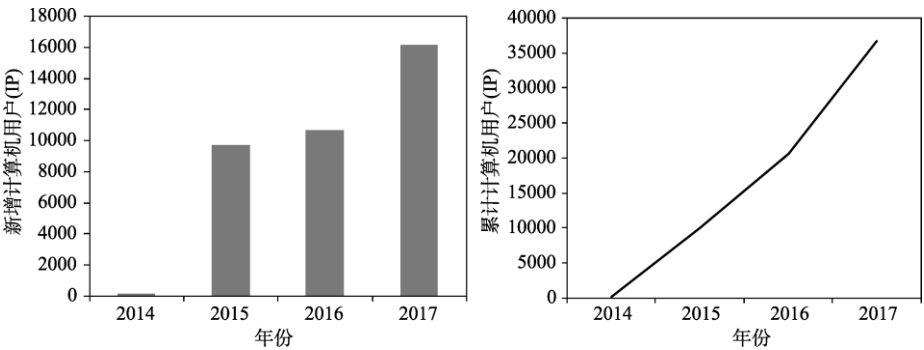


图 1 新增与累计计算机用户（IP）统计图

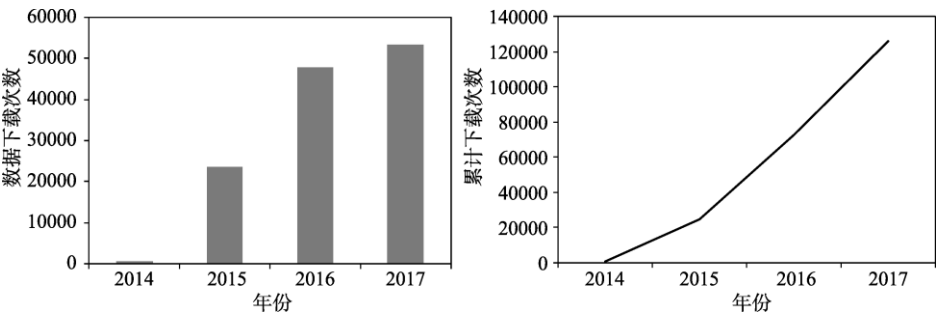


图 2 数据文件下载次数与累计下载次数统计图

6 讨论

毋庸置疑，全球变化科学研究数据的出版，实体数据与数据论文和科学发现论文的关联出版为数据质量控制、数据共享和数据再应用开辟了新途径。由于这是一项新生事物，在数据出版领域还有很多问题有待解决。目前，比较突出的几个问题有：（1）科技界对数据出版问题的重视程度不够；（2）数据知识产权保护力度不够；（3）数据引文混乱；（4）在全球变化科学研究数据覆盖区域方面，亚洲以外的区域数据集相对比较少，（5）国外作者数量比较少。

对于第一个问题，虽然 2017 年中国科学院地理科学与资源研究所、中国地理学会在中国科学院科学传播局的支持下，启动了“全球变化科学研究数据出版与共享”百校传播计划。但是，与大数据发展态势相比还远远不够。科学研究数据出版的全面推动还需要政府主管部门、

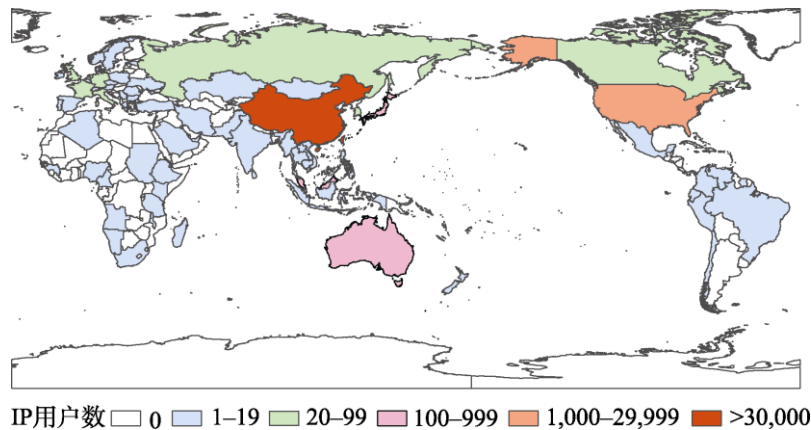


图 3 全球变化科学研究数据计算机 IP 用户世界分布图

各类基金会、科研院所、大学有关部门给予重视，在科学贡献评价体系中给予数据出版成果应有的地位。对于数据的知识产权保护问题，普遍存在很多模糊认识。什么样的数据具有自主知识产权，如何在知识产权保护的同时实现数据共享，这些问题在科技界很多方面还是空白领域，亟待加强。对于科学数据引文问题，由于以前缺少数据引文标准、格式不统一，不规范。“全球变化科学研究数据出版系统”明确要求最终用户在使用数据时，在参考文献中以引文的方式标引数据来源，数据引文包括数据集（实体数据）引文和数据论文引文。对于数据覆盖区域亚洲以外数据集较少、国外作者较少的问题。有必要在以后的工作中，加大宣传力度和政策导向，促进国外学者参与全球变化科学研究数据的出版，鼓励国内外学者合作出版。

参考文献

[1] Future Earth (2013). Future earth initial design: report of the transition team[R]. Paris: International Council for Science (ICSU).

[2] 全球变化研究国家重大科学研究计划专家组编. 中国全球变化战略研究报告[R]. 2009.

[3] USGCRP. The national global change research plan 2012–2021: a strategic plan for the U.S. global change research program [R]. 2012.

[4] 《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.

[5] 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.

[6] 江东, 宋献方, 张国友. 中国科学数据共享新的里程碑[R]. 全球变化数据学报, 2017, 1(2): 246–248. DOI: 10.3974/geodp.2017.02.22.

[7] 刘闯. 论全球变化科学研究数据出版[J]. 地理学报, 2014, 69(增刊): 3–11.

[8] 刘闯, 郭华东, Uhliir, 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.

[9] 石瑞香, 诸云强, 江东等. “全球变化科学研究数据出版系统”被评选为“全国大数据优秀产品、服务及应用方案”优秀案例[R]. 全球变化数据学报, 2017, 1(2): 245. DOI: 10.3974/geodp.2017.02.21.