专题 岩石显微图像

《岩石显微图像专题》卷首语

胡修棉[1*]，侯明才[2,3]，赖文[1]

1. 南京大学地球科学与工程学院，内生金属矿床成矿机制研究国家重点实验室，南京 210023

2. 成都理工大学沉积地质研究院，成都 610059

3. 油气藏地质及开发工程国家重点实验室（成都理工大学），成都 610059

*Email: huxm@nju.edu.cn

数据是科学发现的源泉和基石。现代自然科学是建立在数据基础上的实证科学，离开数据，就谈不上科学的发展。长期以来由于传播媒介和技术的局限性，科学家的研究成果仅能以主要证据和结论的方式在专著、会议论文或期刊论文等载体上发表，而在科学研究过程中产生的大量直接或间接的数据多被埋没，甚至丢弃，这是一种极大的资源浪费。随着科学技术的快速发展，科学大数据呈现爆发式增长[1]。同时，信息技术的飞速发展，存储和传播的成本一再下降，使得数据和信息的存储、传播所所面临的容量桎梏、时间和空间藩篱逐渐打破，科学数据的公开和分享的技术瓶颈被攻克，操作也变得简单易行。

科学数据共享是科技进步的新动力和社会发展的重要需求。科学数据的共享，一方面可以提高研究结果的可检验性和公信力，另一方面能够扩展科学研究的范围，拓宽科学研究的视角，产生更多的科学知识，还可以为科研以外的生产活动进行开发和利用，产生难以预料的社会和经济价值[2]。正是由于科学数据共享的巨大价值，如何保存、利用科学大数据已经成为全世界科学家所关注的热点问题，也得到各国政府、资助机构、出版机构、科研单位、公众等关注和推动。

鉴于地球系统的高度复杂性，其研究方法和指标体系纷繁庞杂，数据以多种形式呈现，如图形图像、文字描述、数据表格等。长期以来，由于缺乏统一、高效的地质数据存储标准和机制，海量的地质数据散布在出版物中，或者分散地储存在研究者手中，不但无法整合利用，甚至面临消失的风险。正如张旗先生和周永章教授[3]所言：“在大数据时代，地质观测、野外考察能否被数据化，非结构化数据能否转变为结构化数据，是地质能否进入大数据时代科学殿堂的关键。”地质大数据具有多源（元）异构、时空相关性、复杂性与模糊性、地质体的全球性与国家利益等鲜明特点[4]。蓬勃发展的大数据科学为典型的数据密集型学科——地质学带来了前所未有的机遇与挑战。

长期以来，科学家对岩石显微图像并没有形成统一的标准，迄今也没有统一规格的岩石显微图像数据库。科学家或者科研团队多根据自己的需要和目标来拍摄少量的图像，置于学术论文中或者存放到网络上，作为科研成果的一部分或者教学材料。随着近年来数字图像技术大发展，大规模拍摄和存储岩石显微图像成为了可能。“深时数字地球”（DDE）是由我国科学家发起的国际大科学计划，以“整合地球演化全球数据、共享全球地学知识”为使命，以推动地球科学研究范式的变革为愿景。出于“抢救”数据、促进数据共享和高效利用的目的，由沉积学工作组和古地理工作组发起，与国内沉积学界、古地理学界同行共同协商，特组织出版《岩石显微图像专题》，面向国内外同仁征集符合标准的岩石显微图像数据成果。专题内容包括且不限于：1）项目产生和获取的未发表的岩石样品显微图像数据集；2）岩石

教学样品显微图像数据集；3）已公开发表过的文章中涉及的岩石样品显微图像数据集；4）团队或个人收集的岩石样本显微图像数据集。

不积跬步，无以至千里。通过本专题的组织和出版，一方面积累一批统一标准的、高质量的岩石显微图像集，另一方面也是探索地学暗数据的发掘与共享模式。我们期望并号召地学界的同仁，用实际行动投入到地质大数据的共享和利用中来。人人分享数据，人人受益数据。用小小的行动来探索地学数据的共享机制，扩大基础研究数据的深化研究与应用。

本专题包括 1 篇沉积岩显微图像数据库标准和 21 篇数据论文。这些数据涵盖了青藏高原、塔里木板块、华南板块、华北板块等太古代以来（约 26 亿年至现代）的 12 类岩石 5286 个岩石样本（图 1，表 1）。每一个样本包括每一个薄片的显微图像以及基本特征描述。本专题含 27 张岩石薄片信息表、46 个压缩包、19 333 张显微照片，数据量超过 110 GB。这些岩石样本涉及超过 62 个岩石单元的 129 个剖面或钻孔，样品分布在中国的 28 个省（区、直辖市）以及捷克波西米亚省（表 2）。

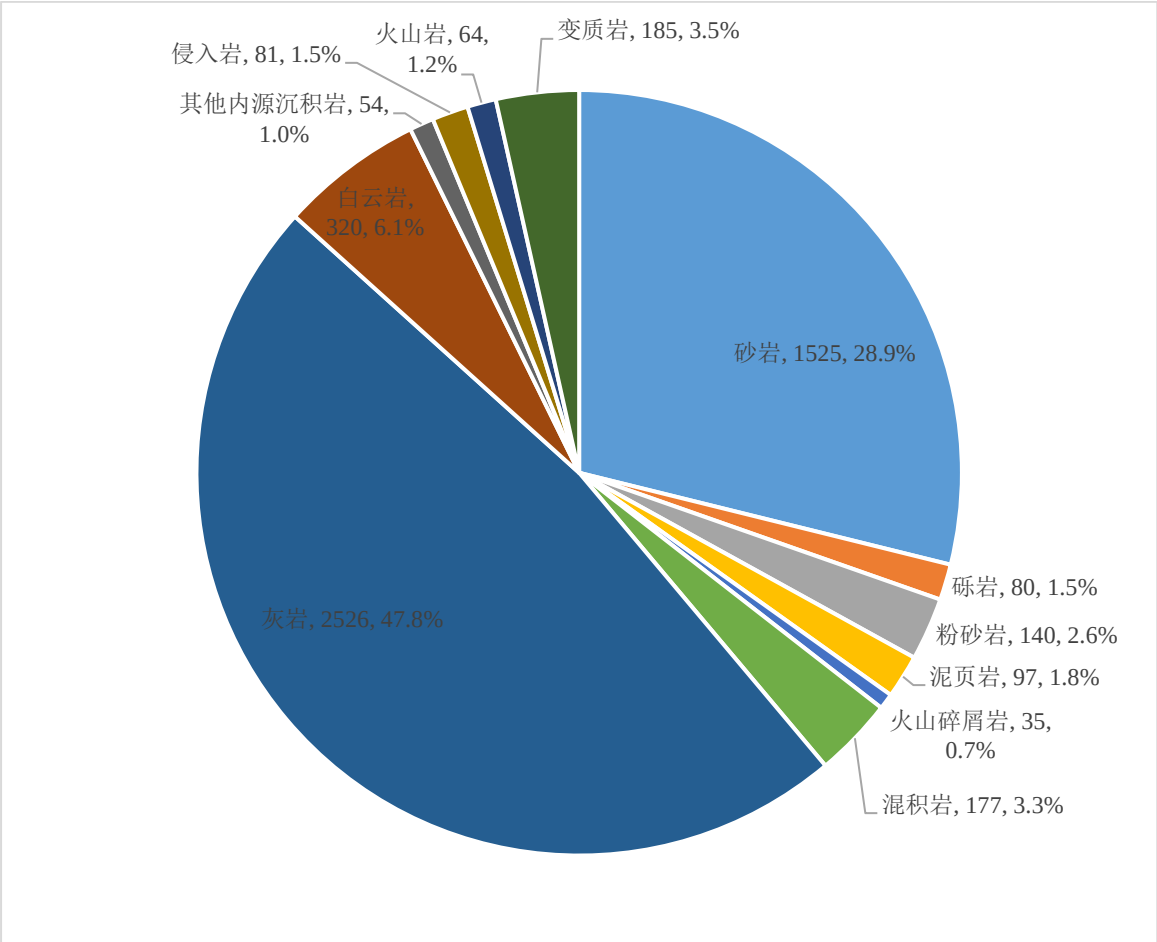

图 1　《岩石显微图像专题》收录的岩石样品类型与数量

表 1　《岩石显微图像专题》收录的岩石样品情况

岩类	砂岩	砾岩	粉砂岩	泥页岩	火山碎屑岩	混积岩	灰岩	白云岩	其他内源沉积岩	侵入岩	火山岩	变质岩	汇总
数量	1525	80	140	97	35	177	2526	320	54	81	64	185	5286

表 2 《岩石显微图像专题》岩石样品地理分布情况

省份	岩石样品数	省份	岩石样品数
西藏	2057	北京	24
新疆	706	甘肃	16
山西	468	广西	16
陕西	395	贵州	16
四川	350	云南	14
河南	193	江西	12
安徽	155	湖南	6
江苏	138	吉林	3
山东	320	辽宁	3
湖北	113	宁夏	3
内蒙古	81	青海	3
浙江	57	台湾	3
重庆	56	天津	3
河北	45	捷克波西米亚省	3
福建	27		

面向大数据时代，本专题数据集建设的目的是便于人机共用，既方便人类读取与使用，也利于计算机去读取与利用。前者很好理解，科学家需要更多的数据集来开展对比研究，来用于教学与大众科普；后者是指伴随着图像技术和人工智能技术的蓬勃发展，基于显微图像的交叉研究成为了可能。实现这种计算机和地质学交叉研究模式的一个重要前提就是要求一定数量的数据集和统一的标准和信息录入格式，这样才能更有效地实现这些数据的整合。从这个意义来讲，本专题数据集丰富了岩石显微图像数据库，一批高质量的图像数据不仅可以满足地学研究人员的需要，也为机器学习与地质学的交叉研究提供了重要的数据样本。此外，这些岩石显微图像不仅可以作为大众科普教育的素材，成为大众了解地球奥秘的窗口之一，也可以直接用于图像密码或图像验证码的素材，而一些独特且具有神秘色彩的显微照片更是具有一定的艺术观赏与猎奇的价值。

感谢《中国科学数据》期刊对《岩石显微图像专题》的支持，特别感谢各数据论文作者的悉心整理和专题编辑的严谨工作。专题得以面世是期刊主编和编委、编辑部、审稿人、作者共同努力的结果。希望岩石显微图像数据集的出版能起到抛砖引玉的作用。衷心希望有更多的地质数据集得以出版，更多的地质暗数据被发掘与共享。只有高质量的地质大数据不断积累，地质学拥抱数据科学才有可能。相信，这一天不会太远。

参考文献

[1] 郭华东. 科学大数据——国家大数据战略的基石[J]. 中国科学院院刊, 2018, 33(8): 768-773.

[2] 郭华东. 问渠哪得清如许, 为有源头活水来——《中国科学数据》发刊词[J/OL].中国科学数据, 2016, 1(1). DOI:10.11922/csdata.0.2016.0014.

[3] 张旗, 周永章. 大数据助地质腾飞: 岩石学报 2018 第 11 期大数据专题“序”[J]. 岩石学报, 2018, 34(11): 3167-3172.

[4] 翟明国, 杨树锋, 陈宁华, 等. 大数据时代: 地质学的挑战与机遇[J]. 中国科学院院刊, 2018, 33(8): 825-831.

论文引用格式

胡修棉, 侯明才, 赖文. 《岩石显微图像专题》卷首语[J/OL]. 中国科学数据, 2020, 5(3). (2020-09-29). DOI: 10.11922/csdata.2020.0088.zh.

附录 1

序号	论文标题	作者	构造分区或地区	样品形成时代	剖面或钻孔数	涉及地层单元数量	样品数/个	显微图片总数/张	鉴定表/张	压缩数据包/个	数据量/GB
1	南京大学岩石教学薄片显微图像数据集	赖文等	中国各地和捷克	未知	—	—	324	2647	3	3	4.95
2	雅鲁藏布江砂粒显微图像数据集	董小龙等	雅鲁藏布江	现代	1	—	2	1876	2	2	10.23
3	藏南特提斯喜马拉雅带晚白垩世–早古近纪碳酸盐岩显微图像数据集	李娟等	特提斯喜马拉雅	晚白垩世–始新世	5	4	465	890	4	4	9.35
4	新疆塔里木盆地西部晚白垩世–始新世岩石薄片偏光显微图像数据集	张世杰等	塔里木盆地西部	晚白垩世–始新世	6	9	682	1364	1	4	2.05
5	北拉萨地体白垩纪郎山组岩石薄片显微图像数据集	许艺炜等	北拉萨地体	白垩纪中期	3	1	559	1134	1	4	25.06
6	藏南日喀则弧前盆地白垩纪陆源碎屑岩显微图像数据集	张艺秋等	日喀则弧前盆地	白垩纪中期	10	3	191	388	1	3	8.9
7	拉萨地体中–北部白垩纪陆源碎屑岩显微图像数据集	赖文等	拉萨地体中北部	白垩纪中期	22	5	402	876	1	3	3.03
8	鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集	晁晖等	鄂尔多斯盆地东北缘	中侏罗世	2	2	78	516	1	2	5.02
9	西藏特提斯喜马拉雅早–中侏罗世岩石薄片偏光显微图像数据集	韩中等	特提斯喜马拉雅	早–中侏罗世	2	3	494	1026	1	2	8.64
10	中扬子区中生代含变质岩屑砂岩的显微图像数据集	马千里等	中扬子地区	三叠纪–侏罗纪	7	6	86	289	1	1	0.81

序号	论文标题	作者	构造分区或地区	样品形成时代	剖面或钻孔数	涉及地层单元数量	样品数/个	显微图片总数/张	鉴定表/张	压缩数据包/个	数据量/GB
11	上扬子西北缘飞仙关组碳酸盐岩显微图像数据集	柴寒冰等	上扬子西北缘	早三叠世	4	1	330	1082	1	1	1.49
12	鄂尔多斯盆地东北缘上古生界盒 8 段砂岩显微图像数据集	史格等	鄂尔多斯盆地东北部	晚二叠世	1	1	280	1144	1	4	3.96
13	中国西南地区二叠纪含火山岩屑砂岩的显微图像数据集	冯薇等	中国西南地区	二叠纪	11	—	43	318	1	1	0.86
14	鄂尔多斯盆地东缘临兴区块上古生界致密砂岩显微图像数据集	李盼盼等	鄂尔多斯盆地东缘临兴区块	石炭纪–二叠纪	24	6	305	660	1	1	5.67
15	南华北石炭–二叠纪太原组灰岩显微图像数据集	马睿等	华北板块南缘	石炭纪–二叠纪	3	1	95	380	1	3	1.98
16	下扬子南部上泥盆统–下石炭统五通群陆源碎屑岩显微图像数据集	蔡文鹏等	下扬子地区	晚泥盆世–早石炭世	12	1	212	856	1	1	3.17
17	塔西北与华南地区晚奥陶世碳酸盐岩（含生物碎屑）显微图像数据集	常晓琳等	塔里木盆地与华南地区	晚奥陶世	2	7	114	348	1	2	1.97
18	鲁西九龙山剖面寒武系苗岭统–芙蓉统之交碳酸盐岩显微图像数据集	辛浩等	华北台地东部	中–晚寒武世	1	2	104	825	1	2	0.32
19	鄂尔多斯盆地中寒武统徐庄组岩石薄片显微图像数据集	钱红杉等	鄂尔多斯盆地周缘	中寒武世	9	1	192	836	1	1	1.98
20	上扬子西北缘震旦系灯影组碳酸盐岩显微图像数据集	齐哲等	上扬子西北缘	震旦纪	2	3	124	1335	1	1	6.74

序号	论文标题	作者	构造分区或地区	样品形成时代	剖面或钻孔数	涉及地层单元数量	样品数/个	显微图片总数/张	鉴定表/张	压缩数据包/个	数据量/GB
21	渤海湾盆地渤中凹陷潜山及上覆砂砾岩显微图像数据集	刘彦鹏等	华北渤海湾	太古代、新生代	2	6	204	543	1	1	4.06
汇总					**129**	**62**	**5286**	**19333**	**27**	**46**	**110.24**

附录 1（续）

序号	论文标题	作者	岩类及数量												汇总/个
			砂岩	砾岩	粉砂岩	泥页岩	火山碎屑岩	混积岩	灰岩	白云岩	其他内源沉积岩	侵入岩	火山岩	变质岩	
1	南京大学岩石教学薄片显微图像数据集	赖文等	18	0	6	12	15	0	15	3	15	75	45	120	324
2	雅鲁藏布江砂粒显微图像数据集	董小龙等	2	0	0	0	0	0	0	0	0	0	0	0	2
3	藏南特提斯喜马拉雅带晚白垩世–早古近纪碳酸盐岩显微图像数据集	李娟等	1	0	0	0	0	3	453	8	0	0	0	0	465
4	新疆塔里木盆地西部晚白垩世–始新世岩石薄片偏光显微图像数据集	张世杰等	54	0	66	9	1	91	436	14	11	0	0	0	682
5	北拉萨地体白垩纪郎山组岩石薄片显微图像数据集	许艺炜等	0	0	0	0	0	0	443	113	0	0	3	0	559
6	藏南日喀则弧前盆地白垩纪陆源碎屑岩显微图像数据集	张艺秋等	159	11	2	5	1	0	1	0	7	1	4	0	191
7	拉萨地体中–北部白垩纪陆源碎屑岩显微图像数据集	赖文等	250	15	26	0	18	2	65	0	6	3	12	5	402

序号	论文标题	作者	岩类及数量												汇总/个
			砂岩	砾岩	粉砂岩	泥页岩	火山碎屑岩	混积岩	灰岩	白云岩	其他内源沉积岩	侵入岩	火山岩	变质岩	
8	鄂尔多斯盆地东北缘中侏罗世碎屑岩显微图像数据集	晁晖等	60	0	8	10	0	0	0	0	0	0	0	0	78
9	西藏特提斯喜马拉雅早–中侏罗世岩石薄片偏光显微图像数据集	韩中等	17	0	0	0	0	70	404	3	0	0	0	0	494
10	中扬子区中生代含变质岩屑砂岩的显微图像数据集	马千里等	83	3	0	0	0	0	0	0	0	0	0	0	86
11	上扬子西北缘飞仙关组碳酸盐岩显微图像数据集	柴寒冰等	0	0	0	0	0	0	295	35	0	0	0	0	330
12	鄂尔多斯盆地东北缘上古生界盒 8 段砂岩显微图像数据集	史格等	280	0	0	0	0	0	0	0	0	0	0	0	280
13	中国西南地区二叠纪含火山岩屑砂岩的显微图像集	冯薇等	43	0	0	0	0	0	0	0	0	0	0	0	43
14	鄂尔多斯盆地东缘临兴区块上古生界致密砂岩显微图像数据集	李盼盼等	305	0	0	0	0	0	0	0	0	0	0	0	305
15	南华北石炭–二叠纪太原组灰岩显微图像数据集	马睿等	0	0	0	0	0	0	95	0	0	0	0	0	95
16	下扬子南部上泥盆统–下石炭统五通群陆源碎屑岩显微图像数据集	蔡文鹏等	134	18	15	44	0	0	0	0	0	0	0	1	212
17	塔西北与华南地区晚奥陶世碳酸盐岩（含生物碎屑）显微图像数据集	常晓琳等	1	0	0	6	0	0	107	0	0	0	0	0	114

序号	论文标题	作者	岩类及数量												汇总/个
			砂岩	砾岩	粉砂岩	泥页岩	火山碎屑岩	混积岩	灰岩	白云岩	其他内源沉积岩	侵入岩	火山岩	变质岩	
18	鲁西九龙山剖面寒武系苗岭统–芙蓉统之交碳酸盐岩显微图像数据集	辛浩等	0	0	0	0	0	0	104	0	0	0	0	0	104
19	鄂尔多斯盆地中寒武统徐庄组岩石薄片显微图像数据集	钱红杉等	18	0	10	10	0	11	108	35	0	0	0	0	192
20	上扬子西北缘震旦系灯影组碳酸盐岩显微图像数据集	齐哲等	0	0	0	0	0	0	0	109	15	0	0	0	124
21	渤海湾盆地渤中凹陷潜山及上覆砂砾岩显微图像数据集	刘彦鹏等	102	33	7	1	0	0	0	0	0	2	0	59	204
汇总			**1525**	**80**	**140**	**97**	**35**	**177**	**2526**	**320**	**54**	**81**	**64**	**185**	**5286**