



中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



科学数据银行
Science Data Bank

中国开放数据白皮书 2023



SPRINGER NATURE



科学数据的开放、共享和应用, 会促进科学界带来新的知识。而大数据、人工智能和大模型的融合发展, 也将激活科学研究的创新力和生命力, 破解更多科学密码。

——陈润生 中国科学院院士 中国科学院生物物理研究所研究员

开放数据状况报告是一项全球调查, 旨在深入了解科研人员对于开放数据的态度和体验。我们很高兴与中国科学院计算机网络信息中心携手, 合作发布一份有关中国开放数据情况的报告, 以便出版机构、科研资助机构和科研机构更好地了解科研人员的看法, 以及需要以哪些支持来帮助他们将数据公开。作为科研界的积极合作伙伴, 施普林格·自然致力于开创数据共享的新方法, 并支持科研人员使数据共享成为新常态。

——STEVEN INCHCOOMBE (史蒂文·印驰库姆) 施普林格·自然
科研市场总裁

中国开放数据白皮书

2023

目 录

前 言	4
2023年中国开放数据情况分析	6
我国科学数据的新型知识产权保护规则	12
中国科学院推动数据开放的政策与实践	14
中国空间科学领域数据开放实践与思考	16
全球微生物大数据共享平台的建设与应用	18
中国对地观测科学数据开放共享的现状和趋势	20
中国农业科学数据开放共享实践与进展	22
基于SI数字框架的计量数据开放共享特点介绍	24
中国科技期刊开放研究数据政策与实践	28
开放数据在中国——施普林格·自然的政策、实践和愿景	31

前言

陈润生 中国科学院生物物理研究所

周园春 中国科学院计算机网络信息中心

开放数据是开放科学的重要组成部分,数据密集型科学发现的新型科研范式正在对科学数据开放共享提出迫切需要。经过中国政府、政策制定者等多方努力,中国的开放数据和开放科学数据事业正迎来蓬勃发展。

中国在全球开放科学的舞台上扮演着重要角色。在2023年的调查问卷反馈中,中国受访者占全球参与人数的11%,排全球第二。同时,中国科研群体对于开放科学有比较高的接受度和支持度。有78%的受访者赞成公开研究数据成为惯例。然而,数据共享在实践中还在敏感数据许可、数据滥用等问题上存在一定阻碍因素。因此,报告对国家政策法规体系、科研评价体系、科研机构的指导作用和数据共享最佳实践提出了相关建议。

同时,报告邀请中国数据开放共享领域的专家共同参与此次报告的撰写。在政策研究与法规层面,来自国家知识产权局的顾昕等介绍了中国对于科学数据这一新型知识产权的保护规则,并对大规模科学数据集提出保护建议,较于个体科学数据已开展的系列知识产权合规工作,大规模数据集还缺乏完善的制度保障,亟待构建一种新型产权制度。

在建设实践层面,来自中国科学院科学数据总中心的陈昕等从数据政策建设实践、科学数据管理与共享服务实践、科学数据治理能力实践、项目数据汇交实践几个方面系统介绍了中国科学院在科学数据开放的政策与实践上的建设情况。

涉及具体学科领域的数据共享工作,报告邀请了5篇来自国家科学数据中心的观点报告。科学数据是重要的国家基础性和战略性资源,而地球科学、生物学、农业农村等领域产生的科学数据,具有极高的经济价值和社会价值,做好这些数据的收集、储存、保护、开发与利用工作,将对我国国民经济和社会发展产生深远影响。微生物和农业等领域均重视数据平台建设,促进领域数据的海量汇聚,形成了有效的整合效应,促进科学数据资源的共享利用。另一方面是数据标准建设达成一定成效,形成数据共享的基础,提升数据质量,实现数据互联互通。

来自国家空间科学数据中心的邹自明等介绍了国内外空间科学领域数据开放发展情况,从领域科学数据的高质量供给、高水平开放共享、开放生态建设等方面具体阐述了中国在推动空间科学领域开放共享过程中的实践与进展,并提出了面向数据与智能融合驱动的新型科研范式持续推进领域科学数据开放与利用的展望。

来自国家微生物科学数据中心的马俊才的介绍,国家微生物科学数据中心以世界微生物数据中心(WDCM)为平台,倡导了全球微生物菌种保藏目录(GCM),在微生物资源共享和挖掘方面建立一套国际标准体系,建立全球权威的微生物组学参考数据库和数据分析平台,有效促进全球微生物数据的汇聚、融合、共享与利用。

国家对地观测科学数据中心的李国庆系统地介绍了中国对地观测数据的特点,并分析了中国对地观测数据开放共享的参与方、开放现状及各方态度;总结了中国对地观测领域开放数据存在的问题及机遇。

国家农业科学数据中心的周国民从分析农业科学数据的特点为入口,全面梳理了国家农业科学数据中心在促进农业科学数据开放共享上的系列举措,分享了相关工作成效,并对未来工作进行了展望。

国家计量科学数据中心的熊行创首先介绍了基于SI数字框架的计量数据开放共享特点,并提出计量可追溯性和机器可读、可理解、可操作性是数字时代计量数据作为质量信任之锚的显著特点。

科学数据在学术出版中起到了保障科研诚信、促进学术交流的重要作用。本次问卷的大部分受访者也赞同这一观点,认为“论文出版是最主要的数据获取方式”。在中国科技期刊的开放数据实践进展上,来自《中国科学数据》的孔丽华研究表明“中国科技期刊卓越行动计划”的资助刊物中有超过81%的“领军期刊”已制定了数据政策,但这在中国科技期刊总量(5000多种)中占比,仍微乎其微。此外,数据论文出版作为一种新型出版模式,正在积极推动数据引用和重用,释放数据更大的价值。中国创办的数据期刊的论文发文量正处于上升期,并开展了数据共享和出版、数据政策建设的丰富实践。

作为出版机构的施普林格·自然介绍了其在中国的数据政策、实践与愿景。施普林格·自然支持中国的开放数据政策和数据存储库,并建议了在中国鼓励数据共享的最佳实践。

我们高兴地看到,本次调查中大部分科研群体赞成开放科学数据成为常态,并且希望得到共享数据能够带来更多的学术认可,这也反映了科研群体对开放科学数据这一概念具有较为正面的认知,并希望以此开展更多学术交流、获得学术认可。当然,政策法规体系建设、科研评价的激励、科研机构的实践指导等,都能够极大帮助科研群体在科学数据开放共享过程中消除障碍、提供便利,逐步形成驱动数据开放的强大动因。我们呼吁,科研机构、高校、学术期刊、出版商、数据存储库、技术提供商等科研共同体,在拥抱开放科学的道路上继续携手前行,共同为构建开放科学和开放数据大家庭努力。



2023年中国开放数据情况分析

姜璐璐 张泽钰 李宗闻

中国科学院计算机网络信息中心

一、报告情况介绍

2023年,是figshare与Springer Nature连续面向全球科学家发放《开放数据状况调查问卷》的第八年。据统计,此次调查中,共收到来自中国的有效问卷反馈642份,中国受访者占全球参与人数的11%,位列第二。本报告以本次调查收到的这642份中国学者的问卷数据为依据,对中国科研群体在开放数据的动机、存在的挑战、看法和实践行为等方面展开调查分析。

根据对中国受访者的数据分析,报告呈现以下要点:

- 有78%的受访者赞成将公开研究数据作为常规惯例;
- 56%的受访者认为其在共享数据的学术认可上“获得太少”,20%回答“不知道”;
- 74%的受访者表示不清楚“数据管理计划”的概念,49%的人表达需要相关培训支持;
- 49%的受访者重复使用过他人或团队公开共享的数据;
- 已发表的论文是最主要的数据获取方式(69%)。

在参与此次调查的中国受访者中,60%来自高校,是占比最高的人群,16%来自医院或医疗组织,14%来自研究机构,6%来自医学院,3%来自企业,另有1%来自政府机构。在学科分布上,从事医学(29%)研究的受访者比例最高,其次是生物(18%),工程(9%)和地球与环境科学(9%)领域。从职业阶段分布上看,约有47%的受访者处于职业早期(于近5年内发表了第一篇论文)。因此,本次调查可能在一定程度上反映出中国青年科研群体对待开放数据的认识。

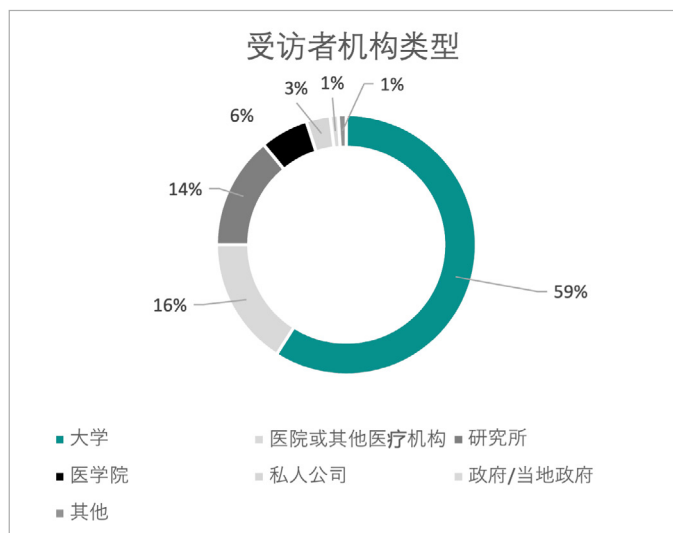


图1 受访者机构类型

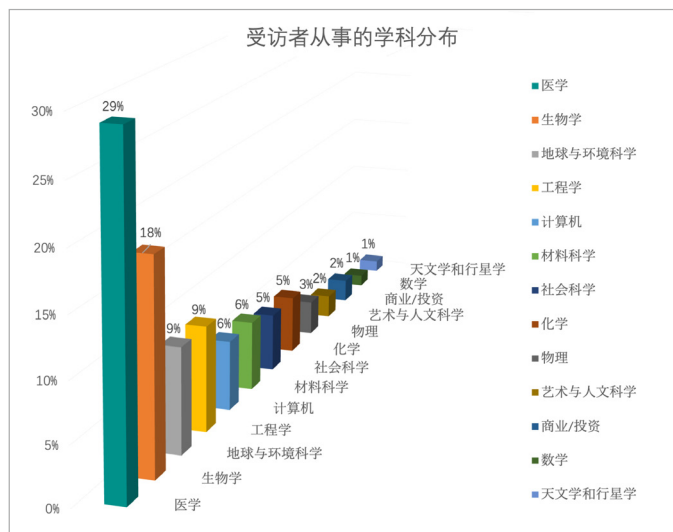


图2 受访者从事的学科分布

78%的受访者赞成将公开研究数据作为常规惯例

二、主要发现

1. 数据共享的驱动与阻碍

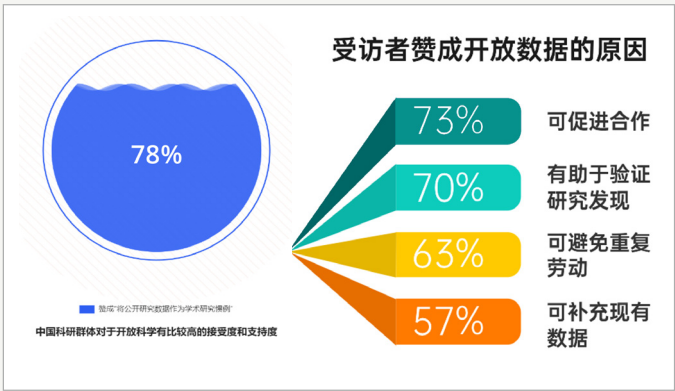


图3 中国科研群体对开放科学的支持度及赞成开放数据的原因

中国科研群体对于开放科学有比较高的接受度和支持度。有78%的受访者赞成将公开研究数据作为学术研究惯例。其中，73%的受访者认为开放数据可促进合作，70%认为“有助于验证我的研究发现”，63%认为“可避免重复劳动”，57%认为“可补充现有数据”，仅1%的受访者从未从数据共享中获益。

50.3%的受访者认为资助方应将研究数据的共享作为获得经费的要求之一（29.4%反对，20.2%回答“不清楚”）；若经费申请人没有遵循项目的强制共享数据要求时，44.9%的人认为申请应遭拒绝或给予其他相应处罚（30.4%反对，24.8%回答“不清楚”）。

数据共享依旧存在诸多阻碍因素。在问及关于数据共享的潜在问题或担忧时，57%的受访者担心数据“包含敏感信息或数据共享前须获研究参与者许可”，医学领域的受访者对此的担忧尤为突出（31.96%）；此外，中国是全球（43%）受访者中对此表达最多顾虑的国家。中国受访者担忧的其他问题还包括“数据滥用（39%）”，“其他实验室抢发研究成果（36%）”以及“不确定数据版权和数据许可（36%）”等；只有约10%的受访者表示对数据共享没有顾虑。

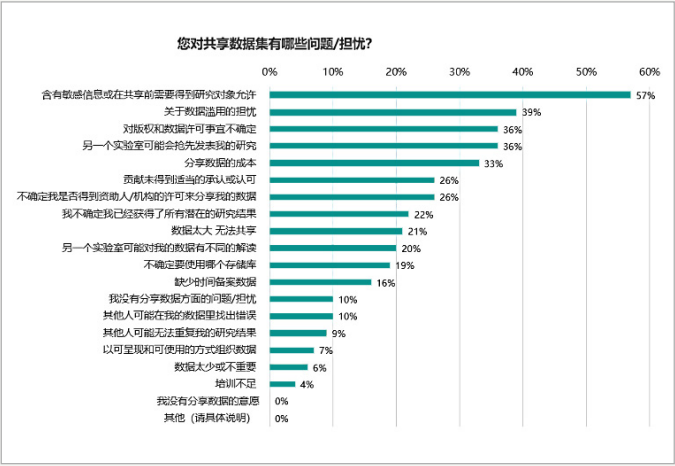


图4 受访者对共享数据的担忧

提升学术认可与影响力是最重要的数据共享驱动因素。调查发现，科研群体最关注的仍是研究成果是否获得认可，及产生的学术影响力。获得“完整的数据引用（69%）”“研究论文被引用（62%）”“提高研究的影响力和知名度（51%）”是中国受访者数据共享的最主要驱动因素。此外，政策要求也发挥着重要的驱动作用，有48%的受访者会因“期刊/出版商的要求”共享数据，38%和36%的人会应资金资助方或所属机构的要求共享数据。

但数据共享尚未获得足够的学术认可。当被问及研究人员目前是否因共享数据而获得足够的承认或认可时，超半数的受访者（56%）选择了“不，他们获得的承认太少”，值得注意的是，较上一年度相比这一比例正在降低。24%的受访者认为数据共享者已获得足够的认可（18%）或过多认可（6%）且这一比例正在上升。

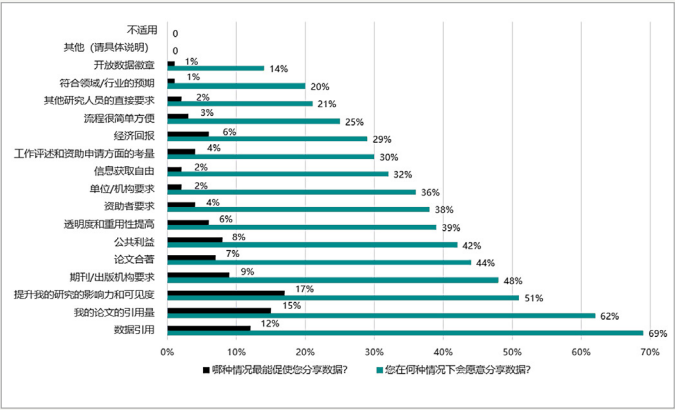


图5 促使受访者共享数据的驱动因素

在认可形式上，论文的完整引用（41%）是研究人员在数据共享上获得最多的认可形式。其次是数据共享者作为论文的合著者（33%），次之是作为工作考核的内容（22%）。但是，有30%的受访者表示从未在数据共享上获得过任何认可，67%的学者表示他们并未因数据共享而有机会加入任何合作项目。

数据共享尚未获得足够的学术认可

在项目进行中，您会将研究数据存储在哪？

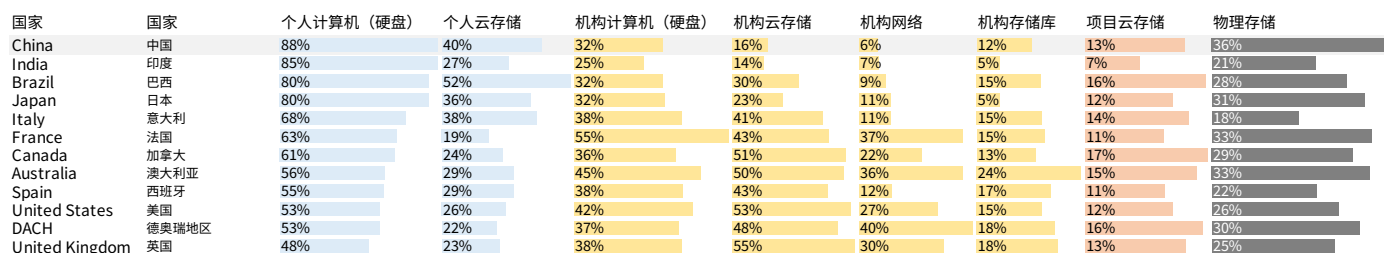


图 8 研究数据的存储方式

2. 科学数据管理的实践

如果您需要制定一个实用的数据管理计划，您认为自己在多大程度上有能力这样做？



图 6 DMP的制定能力评估

中国受访者关于数据管理计划 (Data Management Plan, DMP)

的认知、实践能力及实施支持上普遍需要提升。调查显示，有72%的中国受访者不清楚或者不知道DMP的概念。涉及DMP的制定能力评估，近半（49%）的中国受访者表示自己不具备制定可实施的DMP能力，需要更多培训和信息，仅51%的受访者在不同程度上表示能够胜任。谈及DMP的制定动因上，40%的受访者表示制定数据管理计划是出于行业规范或期望的考虑，39%的人出于项目资助方的要求，36%的人出于机构要求考虑，也有部分人表示会出于个人选择而非其他人要求制定DMP（32%）。

在数据管理计划的实施上，80%的中国受访者有过创建数据管理计划的经历，但实施过程面临诸多挑战。40%的受访者认为缺乏受过训练的专业人员开展工作，39%的人认为实施数据管理计划过于耗时，31%的人认为过于耗资。此外，技术上的挑战诸如数据存储、数据组织等问题也较为突出（37%）。

您认为从数据管理计划中的哪些方面可通过进一步技能培训/信息中受益？

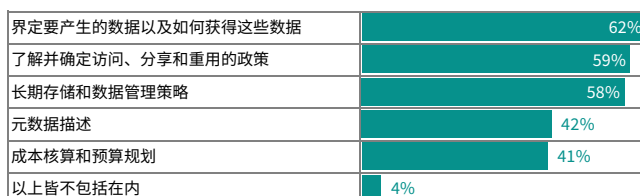


图 7 受访者希望得到的有关数据管理计划的培训内容

相比于全球其他国家和地区，中国受访者在项目进行阶段尤其倾向将数据存储在个人计算机上（88%），40%的人选择使用个人云存储。选择使用机构服务的人群比例明显较低（机构云存储16%，机构存储库12%，机构网络6%）。此外，约有36%的人选择使用物理方式（如纸质笔记本等）保存数据。

较少的受访者（15%）会整理所有收集到的数据以便分享。有34%的受访者表示会整理计划公开的数据，21%的受访者会整理部分数据用于公开，18%的受访者仅对计划分享给同事或他人的数据进行整理。另有约7%的学者表示缺乏相关资源，但乐于开展此项工作。在数据整理的支持工作上，中国学者更倾向通过所在机构获取支持。全球受访者中50%的学者选择从所在机构获取专家帮助，而中国学者中这一比例占到70%，他们通过实验室（46%）、所在部门（35%）、机构图书馆（17%）获得过专家帮助。

较多数中国受访者（56%）倾向在研究发表后再公开自己的数据。此外，14%的学者会选择在提交研究论文时公开数据，只有3%的受访者会在数据收集的当下立即公开数据。另有9%的受访者表示在他人提出要求后才会公开数据，8%的受访者表示不会和主要合作学者以外的人进行数据分享。在共享方式上，大多数中国受访者倾向使用机构存储库（60%）来共享数据，次之是以论文补充材料的形式（46%）共享数据。关于资金来源，36%的受访者共享数据的资金支持来自所属机构，35%的人表示不清楚，27%的人表示更支持有明确数据共享预算的项目。

相当比例的中国受访者有过数据再利用的经历。统计发现，有一半的中国受访者（50%）再利用过自己的数据，49%的人再利用过他人的开放数据，14%的人再利用过非公开的数据用于研究。



图 9 开放数据的获取途径

80%的受访者有过创建数据管理计划的经历，但实施过程面临诸多挑战

已发表的学术论文是最主要的开放数据获取途径 (69%)。其次分别是通过数据仓储库获得数据 (62%) 和通过数据论文获得数据 (46%)。在谈及数据质量的决定因素时, 53% 的中国学者认为“数据是新的 (如最近一年内发布的数据)”是重要的决定因素 (26% 认为“格外重要”, 27% 认为“十分重要”); 此外, “可视化的数据反应了原始数据的真实其情况 (45%)”、“数据与已发表的研究结论一致 (45%)”以及“具备完整的相关元数据 (45%)”也是“格外重要”或“有些重要”的因素。

3. 数据管理的基础支撑需求

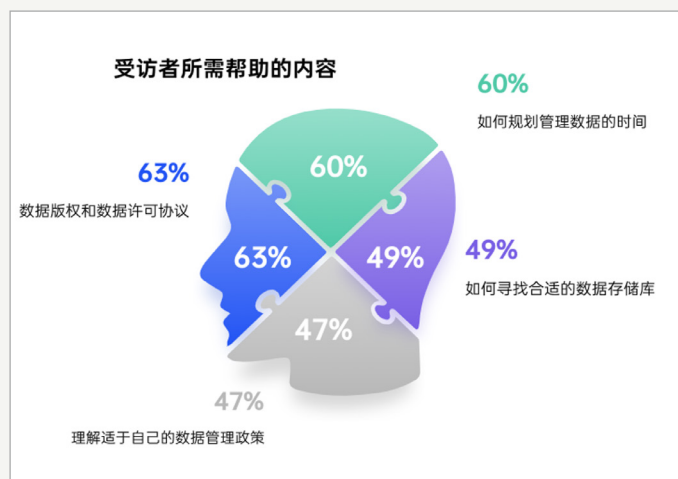


图 10 受访者所需帮助的内容

大多受访者希望得到同事或导师、数据仓储库、机构图书馆和开放数据软件提供商的帮助。关于科研人员管理和共享数据, 过半受访者希望得到同事或导师 (58%)、数据仓储库 (53%)、机构图书馆 (52%) 和开放数据软件提供商 (52%) 的帮助。这与现实情况较为接近, 有超六成 (61%) 的受访者表示已接受过关于如何计划、管理、共享科研数据的相关帮助, 这些帮助主要来自于同事或导师 (63%)、开放数据软件提供商 (58%)、机构图书馆 (57%)。对于科研人员已得到的帮助支持, 有三成左右的受访者认为机构或组织 (29%)、资助方 (37%)、出版商 (32%) 提供了很好的帮助, 20% 左右的受访者得到了一般的帮助。

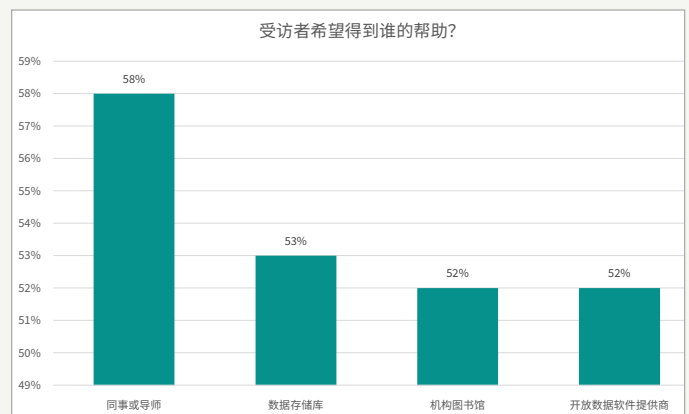


图 11 受访者希望得到谁的帮助

在帮助内容上, 超六成受访者希望了解数据版权和数据许可协议的相关知识 (63%), 以及他们如何规划管理数据的时间 (60%)。还有近半受访者表示需要了解如何寻找合适的数据存储库 (49%)、理解适于自己的数据管理政策 (47%)。

多数受访者认为中国在数据开放共享上给予了较强度支持。

从政策制定上, 接近半数受访者在最近的科研过程中找到了机构 (47%)、出版商 (45%)、资助方 (37%) 以及校方 (36%) 的数据政策。多数受访者希望机构 (67%)、资助方 (58%)、校方 (54%) 和出版商 (52%) 指导他们了解对应的数据政策, 以使研究数据开放可用。多数受访者认为中国对于研究数据开放共享给予了较强度度的支持 (65%)。

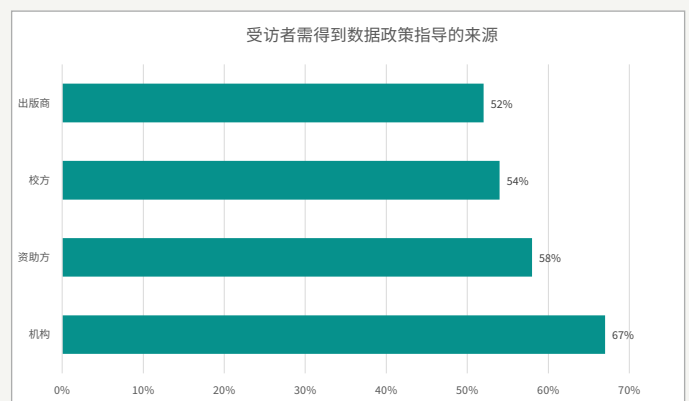


图 12 受访者需得到数据政策指导的来源

此外, 调查数据还显示, 较多的中国数据受访者已在使用人工智能工具辅助他们进行数据收集 (62%)、数据处理 (55%) 和元数据创建 (53%) 工作。

三、主要结论和建议

1. 完善国家政策法规体系为研究人员提供规范指引

《科学数据管理办法》至今已颁布五年,“数据二十条”(《关于构建数据基础制度更好发挥数据要素作用的意见》)已于2022年底发布,但关于科学数据的权属究竟应如何界定和规制,目前仍无明文规定予以落实。实践中,科学数据权益、论文版权、出版转让协议等内容往往相互交织,其中所涉权利纷繁复杂,研究人员在共享数据时不免顾虑。

近六成受访者表示共享数据时担忧数据包含敏感信息或数据共享前须获得研究参与者许可。值得注意的是,受访者中有近五成来自医学(29%)和生物(18%)领域,这些领域的研究活动更多涉及个人信息、人体器官、人类遗传资源等敏感数据,准确厘清数据共享与保护的边界对很多科研人员而言存在诸多困难。此外,36%的受访者表示对数据版权和数据许可协议事宜不太确定,超过六成受访者表达期望了解数据版权和数据许可协议的相关知识。

针对中国开放数据过程中存在的诸多现实困境,建议尽快完善相关政策法规体系,如尽快出台有关科学数据权属、数据共享规范等内容的管理细则,为研究人员合法合规合理的开展科学数据共享活动提供指引。

2. 优化科研评价体系激励研究人员共享数据

提升学术认可与影响力是最重要的数据共享驱动因素,如完整的数据引用(69%)、研究论文被引用(62%)和提高研究的影响力/知名度(51%)。但有三成受访者认为其并未从数据共享中获得任何认可或承认,超五成受访者认为研究人员在数据共享中获得的认可和承认太少了(56%)。因此,数据开放共享需要进一步的激励机制建设:

- (1) 完善配套措施,规范数据引用,在科研共同体内建立数据引用惯例与共识,将数据引用情况纳入学术影响力评价体系,认可数据共享在科研全生命周期的重要性及价值。
- (2) 在科研职业晋升,人才评优评定,项目申请、项目验收结题等过程,认可接收科学数据作为学术成果项并开展评估评价,以激励科研数据共享实践。
- (3) 在学术论文发表过程中,期刊可考虑为共享数据的文章开通绿色通道(如优先评审、优先出版等),并建立论文和关联数据的对应关系,增强论文与数据的连接性。

提升学术认可与影响力是最重要的数据共享驱动因素

3. 科研机构 and 高校应在数据管理实践中有效发挥指导作用

中国受访者关于数据管理计划的认知、实践能力及实施支持上普遍需要提升,在专业训练、耗时耗资、技术难题等实操问题,仍面临诸多困难,也迫切期望获得相关培训和信息。调查发现,所在机构和高校是中国受访者最为希望得到数据管理计划、数据开放共享实践相关帮助的来源。因此,我们的机构、高校以及图书馆应在数据管理实践上发挥更大的作用:

- (1) 机构、高校及图书馆应肩负起普及数据管理和开放共享实践的职责任务,广泛宣传数据管理计划,提升科学数据管理认知与数据开放共享的认同程度;
- (2) 为科研人员提供更全面的数据开放共享实践指导,针对数据管理计划、数据版权、数据许可协议等重要内容开展配套培训,帮助科研人员提升科学数据管理的基础素养;
- (3) 与数据存储库、软件提供商开展深度合作,共同建设数据管理和开放共享的基础设施,为机构内科研人员提供便捷可信的数据出版服务平台和数据传播引用通路。

4. 鼓励数据共享的最佳实践

问卷调研结果显示,有很多中国学者有过重用他人数据的经历。其中,通过已发表的论文而获得数据是最主要的数据获取途径,次之是通过数据存储库和数据论文。同时,大部分的中国学者乐于将其共享的数据服务于重用(包括重复研究、重新分析、重新解读、单独性重用和结合性重用)。因此,项目资助方、期刊、机构等应在科研共同体内广泛鼓励数据共享的最佳实践:

- (1) 鼓励更多的科研工作者在研究论文发表时撰写“数据可用性声明”,提升成果支撑数据的可获取性和研究的透明性。
- (2) 鼓励数据成果作为数据论文的形式进行出版,一方面提升共享的数据的可理解性和可重用性,另一方面可纳入现有科研评价体系获得相应的学术认可。
- (3) 鼓励科研工作者把可共享的科学数据存储专业的数据存储上,确保数据的长期、稳定、可靠存储的同时,在FAIR共享原则框架下有效实现数据共享。

机构、高校以及图书馆应在数据管理实践上发挥更大的作用

我国科学数据的新型知识产权保护规则

顾昕 国家知识产权局知识产权发展研究中心

刘洁 中国科学技术大学知识产权研究院

在开放科学的背景下,科学数据的开放共享具有支撑各领域科学研究、降低研究成本、增强科技创新能力等重要意义。近些年,我国科学数据的开放共享工作取得了显著成效。

科学数据需要开放共享,2018年国务院出台的《科学数据管理办法》规定,对于“政府预算资金资助形成的科学数据应当按照开放为常态、不开放为例外的原则...面向社会和相关部门开放共享”。在各项政策的引导和鼓励下,越来越多的中国学者开始发表数据论文,将优质的科学数据开放全球共享¹。

一、科学数据作为整体数据集合的价值

一方面,基于社会发展对科学数据的巨大需求,科学数据需要开放共享。但另一方面,如何进一步发挥由大量个体数据所形成的整体数据集合的作用,是当前数据时代发展面临的重要议题。与个体数据不同,整体数据是指具有相当数量级的个体数据经一定算法加工处理所形成的数据集合,体现了数据处理者所付出的劳动和投入的资本,典型的例子是人工智能大模型产品,个体科学数据除了作为通用人工智能大模型的优质训练数据之外,往往也汇集形成诸多科学方面的人工智能大模型产品,譬如在算法加持下形成的ChatGPT数据集合,在个体数据提供的信息价值之外,具有全新的功能。《科学数据管理办法》在规范科学数据共享遵循“非营利原则”的同时,也鼓励社会组织和企业等对科学数据进行分析挖掘,形成有价值的科学数据产品,开展市场化增值服务。

二、大规模科学数据集合的保护规则有待完善

随着我国科技、知识产权等法律制度的不断完善和保护意识的不断提高,围绕着个体科学数据的保护开展了一系列合规工作,包括个体数据包含的信息是否侵犯著作权,是否违反个人信息保护规定,如何与相关主体签署合同获得授权等等。但是,相较于对个体数据的重视,科学数据作为整体数据集合发挥作用时,目前并没有十分完善的制度予以保障。

尽管在信息层面上,整体数据结合在满足一定条件时可以沿用现有的法律制度获得保护,譬如作为《著作权法》的汇编作品、利用商业秘密制度,或主张《反不正当竞争法》的一般条款,但对于个体科学数据汇集形成的具有全新功能的大规模数据集合本身应如何保护,目前没有明确的法律规定。

三、构建数据知识产权保护规则的试点探索

国家知识产权局正在探索构建数据知识产权保护规则,尝试建立符合这种全新产权客体的保护方式。自2022年11月开始,确定北京市、上海市、江苏省、浙江省、福建省、山东省、广东省、深圳市等8个地方作为试点地方,在制度构建、登记实践、权益保护、交易使用等方面进行探索。

具体而言,数据知识产权是借鉴知识产权思路所构建的一种新型产权制度,与《数据二十条》²淡化所有权、强调使用权的理念相一致,数据知识产权也是从数据使用权的角度考虑来构建。

1. 保护对象。数据知识产权的保护对象是依法依规获取的、经过一定规则处理的、具有实用价值的数据集合。其中,“依法依规获取”指的是数据知识产权保护对象的安全合规性要求,主要指原始数据的获取应符合《网络安全法》《数据安全法》《个人信息保护法》等相关法律、法规有关规定或者合同约定。“经过一定规则处理的”数据集合可以简单表述为“数据+规则”,是指是将经过数据处理者付出劳动和投入改造后进入经济领域的、已经作为生产要素形态存在的数据作为产权的对象,这就排除了可能包括多方主体的原始数据。这里言及的“规则”在当前的实践中主要体现为算法。

2. 保护方式。考虑到后续潜在竞争者的行为自由,为了最大限度避免干扰数据的正常流转,构建的数据知识产权是一种有限的排他权,仅禁止他人不正当获取以及不正当披露和使用数据集合的行为。

¹中国科协技术协会、国际科学、技术与医学出版商协会联合编写,《中国开放获取出版发展报告2022》,科学出版社出版2023年第1版第47页。

²2022年12月,中共中央、国务院印发《关于构建数据基础制度更好发挥数据要素作用的意见》(简称《数据二十条》)。

四、对大规模科学数据集合的保护建议

符合前述条件的科学数据集合的持有者,对于其在数据收集和处理中所付出的劳动及资本投入,建议可以尝试通过目前八个数据知识产权地方试点进行登记寻求保护。

针对科学数据集合的持有者可能存在的顾虑,以下予以简要解释说明:第一,保护对象不是个体数据,而是整体数据,是经过一定规则处理的具有相当规模的科学数据集合,典型的保护对象是科学领域的人工智能大模型产品。第二,数据登记并不意味着必然选择商业化道路,数据持有者登记后可以选择向使用者收费,也可以选择免费,登记仅仅在一定程度上起到明确整体数据集合权属的作用。第三,数据登记并不妨碍他人正当利用数据的自由。譬如对于某家研发自动驾驶技术的公司,如果是自主合法采集用于人工智能训练的街道数据,即便已经存在类似甚至相同的数据知识产权在先登记,在先登记人也不能禁止他人后续合法的数据利用行为。在先登记人只能禁止他人不正当获取和利用同样或类似数据集合的行为。第四,登记规则设计尽量避免对数据持有者产生不利的市场影响。登记程序并不要求数据持有者提供全部数据集合,譬如北京、浙江等试点的登记办法规定,数据持有者提供数据来源、应用场景、数据特征、存证情况、算法规则、样例数据等内容后即可获得登记,充分考虑了数据持有者在商业秘密上的顾虑。

可以尝试通过数据知识产权地方试点进行登记寻求保护

中国科学院推动数据开放的政策与实践

陈昕 刘宁 周园春

中国科学院科学数据总中心 中国科学院计算机网络信息中心

科学数据已成为科技创新的基础性资源和重要驱动力。中国科学院作为国家战略科技力量主力军,长期高度重视科学数据管理与开放共享工作,自上世纪七十年代末启动科学数据库建设以来,科学数据开放共享程度日益提高。过去一年来,为积极响应联合国《开放科学建议书》和《中华人民共和国科学技术进步法》,充分发挥科学数据价值,中国科学院积极推动科研活动全流程的科学数据管理,在保护科研人员权益的前提下,实现科学数据的长期保存、便捷访问与可重复利用,推动开放数据实践。

一、完善科学数据政策,推动项目数据汇交共享的规模化实施

2019年2月,中国科学院发布了《中国科学院科学数据管理与开放共享办法(试行)》,提出科学数据应按照分等级、可发现、可访问、可重用的原则,适时开放共享,并提出将数据管理计划作为项目立项的必要条件等科技项目数据汇交要求。随后,中国科学院及院属各部门、各单位制定并发布了进一步落实的政策和制度,协同推进科学数据政策制度体系的发展与实践。其中,2022年,《中国科学院战略性先导科技专项管理办法》和《中国科学院战略性先导科技专项科学数据汇交管理实施细则(试行)》相继发布,明确了专项科学数据工作方面的若干要求:1)科学数据管理工作应贯穿专项各个环节,并作为项目阶段性考核和综合绩效评价的重要依据;2)院科学数据中心作为科学数据接收和管理方,为专项提供数据管理方面的专业化技术支撑和服务,并提供相关数据的共享服务;3)全院层面对科学数据汇交、共享的进展等进行统筹管理和动态监督,统筹管理。自此,中国科学院科技项目数据规范化管理进一步强化,项目数据汇交与共享开始规模化实施。

二、科学数据中心提供科学数据管理与共享的专业化服务

为保障科学数据管理与开放共享服务的常态化开展,自2019年起,中国科学院启动了院科学数据中心体系建设,并在2021年底

完成了首批32家院科学数据中心的认定,逐步发展地域分布、资源协同的科学数据中心服务网络。科学数据中心作为专业机构,其基础设施环境、科学数据治理人才队伍、以及数据共享的技术平台,为推动项目数据开放共享发挥了非常重要的作用。在项目科学数据汇交与管理工作中,院科学数据中心除提供科学数据存储服务外,还提供了专业化的指导与技术支持,包括:在项目立项初期,指导科研人员形成科学数据管理的工作方案——科学数据管理计划;在实施阶段,支持科学数据的汇交,审核数据质量;按照协议对数据整编与开放共享。

三、发展面向开放科学的科学数据治理能力

中国科学院也特别关注了面向开放科学的科学数据治理能力提升,形成了政策、标准、软件工具三位一体的开放数据解决方案。发展机器可操作的标准规范,重点规范了数据共享中的访问权限、授权许可、数据引用等要求,发布了具有开放兼容性的元数据词表,构成了科学数据开放共享的机制基础。研发的覆盖科学数据全生命周期管理的科学数据中心软件栈,将数据政策和标准要求融入其中,为科研团队和研究机构提供了低时间成本和经济成本的开放科学综合服务。通过开放的科学数据互操作协议,实现了科学数据中心资源的互联互通,在线可共享资源量超过了45 PB。

四、统筹服务的项目数据汇交实践

面向项目数据汇交的业务需求,建设了院项目科学数据汇交管理平台(<https://dmp.casdc.cn/>),依托科学数据中心网络的服务能力,形成了全院项目数据管理的统一入口。平台不仅支持科学数据管理计划的制定,还通过同科学数据中心的互联,对科学数据管理计划的执行情况进行追踪;通过基于科技资源标识的增值服务,还可对科学数据的引用等进行进一步的追踪,实现对科研全流程和科学数据全生命周期的全覆盖。

截至2023年8月底,已有30个战略性先导科技专项通过该平台开展了科学数据管理计划的编制,其中13个专项已开始汇交数据,汇交的资源总量达到102.05TB。这些汇交的资源,按照“开放为常态,不开放为例”的原则,除少数需质控加工后上线外,在审核通过后立即按作者设定的策略在各科学数据中心网站发布共享,已汇交数据开放性达87.2%(完全公开56.5%,保护期10.6%,有条件公开20%)。在此过程中,吸引了2039名科学家个人或团队熟悉开放科学技能,成为数据作者,覆盖164家研究机构,初步形成由一线科学家组成的高水平数据作者团队,为形成高度活跃和可持续发展的科学数据共享生态奠定基础。

五、进一步发展面向开放科学的科学数据治理能力

过去的一年,项目数据汇交工作的推进为全院科学数据规范化管理水平的提升和开放科学文化的建立奠定了重要基础。同时,科学数据银行(Science Data Bank)通过领域和专题社区的建立,探索同科研社区的深入连接,在服务科研共同体方面做出了积极的实践。

未来,将进一步加强同科研创新的联系,以促进开放科学和科技创新为目标,推动科技项目、大科学装置、野外台站等多元场景科学数据的规范化管理与共享服务,提升技术赋能的科学数据治理能力,推动科学数据引用文化和新型科研评价机制的形成,发展开放科学学术环境。

协同的科学数据中心服务网络为项目科学数据的管理与开放共享发挥了重要作用。一线科研人员的参与,为开放数据生态奠定基础。

中国空间科学领域数据开放实践与思考

邹自明 胥鑫 胡晓彦 佟继周 许琦 汤惟玮

中国科学院国家空间科学中心 国家空间科学数据中心

本文介绍了国内外空间科学领域数据开放发展情况,从领域科学数据的高质量供给、高水平开放共享、开放生态建设等方面具体阐述了中国在推动空间科学领域开放共享过程中的实践与进展,并提出了面向数据与智能融合驱动的新型科研范式持续推进领域科学数据开放与利用的展望。

一、引言

人类对太空的探索是一个永恒的话题,从古代诗人屈原在《天问》中提出的“日月安属?列星安陈?”的好奇,到现代空间科学的蓬勃发展,人类一直渴望揭示宇宙的奥秘和探索未知的领域。在漫长的空间探索过程中,源源不断产生的空间科学数据成为了全人类共有的宝贵财富。

加强空间科学领域的国际合作与数据开放共享有助于推动空间科学取得新的突破和更大的进步。尽管面临着国际合作形势、隐私安全与知识产权保护等挑战,空间科学领域的国际主要机构与组织长期以来仍为空间科学数据开放作出了诸多努力。例如,国际空间研究委员会(Committee on Space Research, COSPAR)自1958年成立以来,致力于加强国际空间合作开放,推动全球空间科学研究的发展。由美国地球物理学会(AGU)主导发起的地球和空间科学数据出版联盟(COPDESS¹)于2014年联合领域相关基金会、学术出版机构、仓储库、科研社区组织和研究人员,发起了关于支持和遵循开放数据最佳实践的倡议——《地球、空间与环境科学承诺声明》。

2021年联合国教科文组织(UNESCO)发布《开放科学建议书》²,标志着国际社会推动开放科学的努力进入新阶段,也为空间科学领域的数据开放与利用带来了新机遇。世界主要空间强国均启动部署了开放科学相关战略规划与项目,如美国国家航空航天局NASA推出了开源科学倡议(Open Science Initiative, OSSI³)以及向开放科学转型项目(Transition to Open Science, TOPS⁴),提出将2023年定为开放科学年;加拿大航天局(CSA)发布了2021-2024开放科学行动计划⁵,制定了开放数据、开放获取出版物、利益相关方参与等3个关键领域实施的优先级别和策略;欧洲航天局(ESA)发布《数据与信息开放获取的政策》⁶等。这些举措旨在促进空间科学数据应用与创新生态的健康发展,以加速空间科学进步,寻求重大科学发现。

二、中国空间科学数据开放实践

经过多年努力,中国空间科学数据开放与利用取得了显著进展,具体实践着力点包括发展提供高质量的数据供给,积极推动高水平的数据开放出版,以及营造构建开放合作的数据创新应用生态等方面。这些举措极大促进了中国空间科学数据价值发挥,为推动突破性的科学发现和服务公共利益提供了有力支撑。其中,国家空间科学数据中心(以下简称“数据中心”)作为2019年6月中国首批成立的国家科学数据中心⁷之一,以及目前空间科学领域唯一的国家级科学数据中心,在空间科学数据的治理、开放与利用中发挥了国家平台的积极作用。

1. 高质量数据供给

高质量的数据供给首先需要推进空间科学数据的有效汇聚。数据中心推动了子午工程、空间科学先导专项、月球与深空探测等中国空间科学重大任务产生的科学数据汇交,积极整合国家重点研发计划、国家自然科学基金等百余个不同类型的科研项目产生的数据资源,并通过国际观测网络合作及数据镜像站建设等方式与英国、美国、加拿大、日本、巴西等十多个国家开展空间探测数据交换。这些举措使空间科学数据逐步脱离了分散孤立的状态,加强了数据资源之间的互联互通。

高质量的数据供给同时也需要强化空间科学数据的有序治理。高水平治理离不开标准规范的基础支持,在面向科技资源管理的科技平台标准体系⁸基础上,数据中心团队进一步发展了通用科学数据标准体系及面向空间科学领域的科学数据标准体系,并针对科学数据治理实践域中的关键环节开展标准研究,已相继立项了一批国家标准如《空间科学数据元数据》和《科研项目数据管理指南》等,立项并批准了若干团体标准如《空间环境数据安全分级指南》等。依据相关标准规范形成的空间科学数据全生命周期管理标准化流程,有效促进了领域数据的质量提升、有序管理与安全保存。

2. 高水平开放出版

在国家和主管部门相关办法^{9,10,11}指导和相关国际倡议推动下,中国空间科学领域以“开放为常态,不开放为例外”的原则,尽可能地促进科学数据的公益性开放,同时积极推动领域数据开放从传统数据发布向规范数据出版模式转变,涉及基于数据仓储平台出版、

学术论文关联数据出版和数据论文出版等多种模式，以促进空间科学数据的可发现、可访问、可操作和可重用 (FAIR)，提升研究成果的价值和影响力¹²。

在重大任务数据开放共享方面，推动了一大批中国空间科学任务数据基于数据中心仓储平台面向全球范围开放获取，如“嫦娥”系列任务产品级科学数据、“天问一号”部分2级及以上科学数据、“悟空”卫星伽马光子数据、“慧眼”卫星提案观测数据、“极目”卫星伽马射线暴数据、“夸父一号”首批科学数据等，受到了国内外空间科学领域科研共同体的关注。

在研究者科研数据出版方面，数据中心与科学数据银行开展战略合作，建立了空间科学领域社区ScienceDB•Space，面向科研人员、科研期刊等利益相关者，提供论文关联数据的提交、审核、保存、出版、共享和获取等服务。ScienceDB•Space受到美国地球物理学会 (AGU)、施普林格·自然 (Springer Nature) 科学出版社等国际主流学术出版集团认可，其出版数据集被数据引用索引 (Data Citation Index)、谷歌数据集搜索 (Google Dataset Search) 等多个国际数据索引库收录。

在数据论文出版方面，数据中心与子午工程联合，在《中国科学数据》期刊推出的子午工程数据专刊，对子午工程运行多年来积累的数据产品体系与其中的优质数据集进行了系统介绍。这是中国空间科学领域内首次专业数据期刊以专刊形式出版数据论文，是拓展重大任务数据开放共享方式、提高数据成果价值认可的有益尝试。

基于在领域数据仓储管理与开放出版方面的积极努力，中国国家空间科学数据中心入选了AGU发布的“领域·学科仓储库推荐名单”，成为AGU旗下多期刊推荐的全球21个仓储库之一。中国的空间科学团队也在全球范围内逐步从“数据使用者”转变为更多承担“数据贡献者”的重要角色。

3. 开放合作生态建设

开放合作的数据创新应用生态对于产生科学数据价值的辐射与倍增起到重要作用。中国空间科学领域通过跨领域合作促进交叉融合应用，以更好地解决空间科学及其邻近领域的重大科学问题与公共议题。例如国家空间科学数据中心与国家高能物理科学数据中心、国家天文科学数据中心开展战略合作，在数据技术、数据融合、数据安全、人才培养等方面开展深入合作，并尝试探索联合主题数据目录发布等方式促进跨领域数据发现与应用。

同时，中国空间科学领域的数据与研究团队也正逐步深度参与到各项国际合作与国际组织活动。国家空间科学数据中心作为世界数据系统 (World Data System, WDS) 正式成员，获得CoreTrustSeal (CTS) 国际资质认证，联合签署了COPDESS《地球、空间与环境科学承诺声明》，其团队成员也在WDS、RDA以及CODATA&GOSC等国际合作组织工作组中发挥了积极的作用。

三、未来发展的思考

随着人工智能技术的日新月异发展和以大数据为核心的新一代数字技术融合发展，数据与智能融合驱动的科研范式正在发挥新型创新引擎作用，空间科学领域同样正处于科研范式变革的关键时期¹³。新范式对空间科学数据开放与利用提出了新需求。未来，中国的空间科学数据工作者将在提升数据FAIR化水平的基础上积极发展AI-ready的空间科学数据，着力研发领域专用的AI模型与数据应用工具，建设集大数据、大模型、强算力于一体的空间科学领域开放研究平台，大力推动开放数据、开放模型、开放工具和开放设施相结合的开放共享氛围，形成面向新型科研范式的体系化数据服务能力。

参考文献

¹COPDESS. Commitment statement in the Earth, space, and environmental sciences [EB/OL]. <https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences>. [2023-09-12]

²UNESCO. UNESCO Recommendation on Open Science [EB/OL]. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>. [2023-09-12]

³NASA. Open-Source Science Initiative [EB/OL]. <https://science.nasa.gov/open-science-overview> [2023-09-12]

⁴NASA. SDMWG Strategy_Final-v4 (nasa.gov) [EB/OL]. https://science.nasa.gov/science-pink/s3fs-public/atoms/files/SDMWG_Full%20Document_v3.pdf. [2023-09-12]

⁵Canadian Space Agency. 2021-2024 Open Science Action Plan [EB/OL]. <https://www.asc-csa.gc.ca/eng/publications/open-science-action-plan-2021-2024.asp>. [2023-09-12]

⁶ESA - ESA affirms Open Access policy for images, videos and data. [EB/OL]. https://www.esa.int/About_Us/Digital_Agenda/ESA_affirms_Open_Access_policy_for_images_videos_and_data [2023-09-12]

⁷科技部, 财政部. 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 [EB/OL]. (2019-06-05) [2023-09-12]. http://www.most.gov.cn/xgk/xinxifenlei/fdzdgknr/qtwj/qtwj2019/201906/t20190610_147031.html.

⁸许东惠, 赫运涛, 王志强等. 面向科技资源管理的科技平台标准体系研究 [J]. 中国科技资源导刊, 2020 (2): 1-6, 16.

⁹科技部, 财政部. 科技部财政部关于印发《国家科技资源共享服务平台管理办法》的通知 [EB/OL]. (2018-02-13) [2023-09-12]. http://www.most.gov.cn/xgk/xinxifenlei/fdzdgknr/fgzcgfwj/gfwj2018/201802/t20180224_138207.html.

¹⁰国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知 [EB/OL]. (2018-02-17) [2023-09-12]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

¹¹《中国科学院科学数据管理与开放共享办法(试行)》印发 [EB/OL]. (2012-02-11) [2023-09-12]. https://www.cas.cn/sygz/201902/t20190221_4679910.shtml.

¹²袁雅琴, 胡晓彦, 佟继周等. 大数据开放背景下的我国空间科学数据出版实践 [J]. 中国科技资源导刊, 2022 (1): 89-96.

¹³胡晓彦, 徐寄遥, 邹自明. “大数据&人工智能”驱动的空间天气科研范式变革初步探索 [J]. 数据与计算发展前沿, 2023, 5(02): 24-36.

全球微生物大数据共享平台的建设与应用

马俊才 中国科学院微生物研究所

国家微生物科学数据中心以世界微生物数据中心 (the World Data Center of Microorganisms, WDCM) 为平台, 倡导全球微生物菌种保藏目录 (Global Catalogue of Microorganisms, GCM), 对全球主要微生物资源保藏中心的目录进行标准化整理, 搜集、清理并整合全球微生物领域的微生物资源、组学、文献、专利等数据进行挖掘利用, 在微生物资源共享和挖掘方面建立一套国际标准体系, 建设全球权威的微生物组学参考数据库和数据分析平台。该计划的全球推广, 有效促进了全球微生物数据的汇聚、融合、共享与利用。

一、共同构建全球微生物资源大数据平台, 建立国际合作网络, 促进全球微生物数据的汇聚

微生物资源是生态系统的基本组成部分, 是人类生产、生活资料的基本来源和人类赖以生存的重要物质基础, 是生物技术和产业发展的重要基石。对生物资源的收集、保藏、保护、开发与利用是保障人民生命健康、支撑国民经济可持续发展的重要基础。而微生物资源目录是科学家和产业用户从各国微生物资源保藏机构获取微生物资源的重要有效途径。

世界微生物数据中心 (the World Data Center of Microorganisms, WDCM) 由世界菌种保藏联盟在上世纪60年代建立, 是全球微生物领域最重要的实物资源数据平台。2010年, WDCM落户中国科学院微生物研究所, 这是我国生命科学领域的第一个世界数据中心。国家微生物科学数据中心作为其依托单位, 有效利用区块链技术、生物信息技术、云服务平台和超算资源, 牵头搭建了全球微生物资源数据共享平台。该平台是由中国引领, 52个国家和地区的151家机构参与构建的全球微生物大数据平台基础设施, 平台92%的数据来自于国际合作伙伴, 在国际微生物数据领域产生良好的“全球虹吸效应”, 使大量的全球微生物资源数据系统性地向我国的数据平台聚集, 对全球微生物模式菌株资源进行了有效整合。平台有效促进了全球微生物资源的共享利用, 进一步拓展和深化网络空间国际交流与合作, 共筑全球微生物领域网络空间命运共同体, 促进全球开放科学、经济发展与产业进步。

二、为中国牵头的全球国际合作计划提供数据平台支撑, 促进全球微生物实物资源与数据融合

生物资源是人类社会经济发展和技术进步的重要物质基础, 生物资源有效挖掘利用也是世界各国科技竞争的重要体现。微生物研究所以WDCM为平台, 坚持开展“以我为主”的国际合作, 倡导全球微生物菌种保藏目录 (Global Catalogue of Microorganisms, GCM), 发起全球模式微生物基因组测序计划 (GCM 2.0), 对全球主要微生物资源保藏中心的目录进行标准化整理, 进而从中选择目前未进行测序的模式微生物菌株, 5年内完成超过10,000种的细菌、真菌、古菌模式菌株基因组测序。作为中国牵头的国际大科学计划, 该计划将覆盖全球主要合作伙伴, 帮助解决领域基础和前沿的重大科学问题, 促进微生物分类领域全球开放科学的发展, 也为全球一站式微生物鉴定和认知平台的建立提供重要的数据支撑。全球微生物资源数据共享平台作为国际合作计划支撑平台, 已汇聚超过52万株微生物实物资源的信息, 对全球微生物模式菌株资源进行了有效整合, 有效促进了全球微生物资源的共享。

三、打造国际ISO数据标准, 提升全球微生物高质量基础设施建设, 实现全球数据的互联互通

数据标准是全球数据共享的基础, 是提升数据质量的关键, 是建立国际一流数据库的前提。为了打造高质量国际性数据平台, 中国科学院微生物研究所国家微生物科学数据中心与国家科技基础条件平台中心等单位联合美国、日本、俄罗斯、韩国等九个国家, 共同制定了ISO 21710:2020 Biotechnology — Specification on data management and publication in microbial resource centers (微生物资源中心数据管理和数据发布规范)。该标准于2020年11月9日, 由国际标准化组织生物技术委员会 (ISO/TC 276) 正式发布。这是国际微生物领域的第一个ISO级别的数据标准, 也是我国在国际生物技术标准委员会 (ISO TC276) 主导制定的第一个国际标准。该标准由中国科学院微生物研究所、国家微生物科学数据中心马俊才团队牵头起草。

本标准提供了一组供数据发布的数据字段集,旨在通过应用唯一标识符和统一的数据形式提高微生物资源中心(MRC)在线目录间的数据交换,并有助于通过共享微生物资源来促进实现后续惠益共享。本标准还规定了数据管理和内部数据质量控制的要求,以提高MRC以文件形式记录的数据和信息整体准确性和可靠性,这是高效数据共享和交换的基础。本标准规定了MRC数据管理和发布的要求,包括实现一致格式化的数据格式和提高数据整体质量的质量控制工作流程。这将对数据访问、获取、认证、保存、存储、分发和处置等程序产生影响。还为MRC提供了建议,以促进与微生物资源相关的数据共享和数据整合。标准适用于MRC、监管机构、认证机构组织以及使用同行评估的方案,以确认或认可微生物资源中心数据发布和共享中的数据管理的能力。

参考文献

¹Wu L, Ma J. The Global Catalogue of Microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation [J].Int J Syst Evol Microbiol, 2019, DOI 10.1099/ijsem.0.003276

²Wu L, Sun Q, Ma J. World data centre for microorganisms: an information infrastructure to explore and utilize preserved microbial strains worldwide [J]. Nucleic Acids Research. 2017, 45(D1):D611-D618.

中国对地观测科学数据开放共享的现状和趋势

李国庆 陈祖刚

中国科学院空天信息创新研究院

对地观测数据是人类借助航天、航空以及地面观测平台实施对地球不间断地观测和监测产生的原始数据, 以及通过信息处理再现和客观反映出和人类活动相关圈层的地物状况、地理过程, 确定地学现象的内容、过程及其空间定位和分布的衍生数据。

经过几十年的发展, 中国对地观测已具备面向近地空间、大气、海洋和陆地的高精度、高时空分辨率的数据获取能力。据统计, 自1970年4月24日发射第一颗人造地球卫星“东方红一号”以来, 中国在轨运行的卫星已超过600颗, 其中对地观测卫星有约200颗, 目前已形成了“风云”“海洋”“资源”“高分”“遥感”等多个对地观测卫星体系。

对地观测数据不仅在科学研究、生态环境、农业、土地资源、自然灾害、健康、能源、气候、天气和重大工程的监测与评估等方面得到广泛应用, 而且在数字地球、智慧城市建设中也发挥着重要作用, 并逐步深入到大众生活, 产生了巨大的经济价值和社会价值。对地观测数据已成为国家基础性和战略性资源, 在国民经济、社会发展和国家安全中发挥着不可或缺的作用。

对地观测数据具有海量、多源、多时相、高价值、异构、多尺度、非平稳等特征, 应用于科学研究的对地观测数据被称为对地观测科学数据。中国政府和学术机构历来重视对地观测科学数据的开放共享, 2019年由国家财政部和科技部成立国家对地观测科学数据中心负责专门统筹协调中国对地观测科学数据的开放共享。

一、我国对地观测科学数据开放共享的相关参与方

中国对地观测科学数据的生态链涉及数据生产者、数据管理者、数据分发者、数据使用者等。数据生产者包含对地观测仪器平台(例如卫星)制造商、卫星发射机构、卫星数据接收和处理机构、对地观测衍生数据生产的科学家和工程师等; 数据管理者主要是具有数据所有权的机构及其代理机构, 例如管理公益性陆地卫星数据的中国资源卫星应用中心、公益性海洋卫星数据的国家卫星海洋应用中心、公益性气象卫星数据的国家卫星气象应用中心等; 数据分发者主要是从事对地观测数据销售、共享、传输等活动的机构, 例如各个代理

国内外卫星数据销售的商业公司, 一些具有对地观测科学数据分发权的科学数据中心, 具有对地观测科学数据仓储或者出版业务的期刊出版商、提供卫星数据在线分析或云计算服务的商业公司、企事业单位等; 数据使用者包含科学研究人员、工程技术人员。

二、我国对地观测科学数据开放共享的现状

据统计, 中国对地观测科学数据相关参与方所在的机构超过7万家, 生产和管理的数据总规模接近150PB。这些数据主要以在线形式和离线形式存储, 其中在线存储的数据总量超过95PB, 离线存储的数据规模接近55PB。中国有50%的对地观测科学数据机构提供数据免费开放共享服务, 免费开放共享的数据资源总量占数据总量的47%。中国对地观测科学数据的总用户数超过53万人, 其中在线注册认证登记的用户数超过47万人, 使用对地观测科学数据的机构超过12万家, 总数据使用量超过293PB, 其中以直接下载的方式使用的数据量为149PB。

三、对地观测领域各主体对开放科学数据的态度

我们把对地观测领域的主体按照行业划分为对地观测企业工作人员、行业应用工作人员、科学研究人员, 分别调研他们对地观测科学数据开放共享的态度。结果发现, 科学研究人员认为对地观测科学数据开放共享非常必要, 有95%的人员希望进一步加强对地观测领域科学数据的开放共享。同时有63%的科学研究人员在保证合法权益的前提下愿意共享自己的科学数据给其他人。其次是行业应用工作人员, 有60%的人员希望扩大对地观测科学数据的开放共享, 但他们大多数表示自己所掌握的对地观测科学数据所有权不归自己所有, 不能决定共享给其他人员。最后是对地观测企业工作人员, 他们大多数不希望数据免费公开共享给别人, 除非公开共享数据能使他们获取诸如企业声誉的提升, 产品得到广泛宣传等。总体来看, 科学数据开放共享最相关的群体——科学家群体, 非常支持和赞同对地观测科学数据开放共享, 并且愿意付诸实际的行动。

四、我国对地观测领域开放科学数据的机遇与挑战

由于对地观测科学数据具有巨大商业价值,并且涉及众多类型的主体,中国对地观测领域开放数据既面临重大的机遇也面临一些挑战。

我们面临的机遇有:第一,在科学界,对地观测科学数据开放共享已经是科学家的共识和需求,有较多的科学家愿意积极参与其中;第二,中国有巨量的对地观测科学数据生产者和从业机构,能保证对地观测科学数据资源的充足供应;第三,政府和管理机构积极支持和促进对地观测科学数据的开放共享;第四,对地观测科学数据治理和管理技术能力不断提升,能不断地采取新的技术手段促进对地观测科学数据的开放共享。例如,国家对地观测科学数据中心利用云计算资源,搭建对地观测知识枢纽,实现海量对地观测科学数据按需在线分析,既免除了科学家搬运巨量数据资源的不便,又保证了一些按照管理政策要求不能分发的数据资源的被使用。

我们面临的挑战有:第一,在对地观测领域科学数据开放共享的背景下,如何平衡商业利益和科学家需求。只有实现不同群体间利益的平衡才能进一步促进对地观测科学数据的开放共享。第二,对地观测领域的科学数据资源越来越多,通过网络传输、离线使用的方式越来越满足不了科学家分析使用的需求,如何构建新型对地观测科学数据开放共享基础设施环境和模式是未来面临的巨大挑战之一。第三,对地观测领域开放共享的科学数据资源越来越多,这些科学数据资源的质量如何评估,所有者的权益和贡献如何体现和保证,数据如何被规范引用?进一步加强对地观测领域科学数据开放共享的治理技术和政策研究也是未来面临的重要挑战。

中国农业科学数据开放共享实践与进展

周国民

中国农业科学院农田灌溉研究所 国家农业科学数据中心

一、农业科学数据的特点

农业科学数据是从事农业科技活动所产生的基本数据,以及按照不同需求而系统加工整理的数据产品和相关信息,是农业科研活动的基本产出之一,也是一种特殊形式的农业科研成果。农业科学数据是推动农业科研向深度和广度发展的最为活跃、最为基础的科研要素。

农业科学数据主要有三个来源:一是农业科技基础数据库群,2002年科技部启动“国家科学数据共享工程”,2012年科技部启动“科技条件平台建设”,农业领域通过“农业科技基础数据库建设与共享”“农业科学数据共享中心”等科技部的平台条件项目支持,围绕作物科学、动物科学和动物医学、农业资源与环境、草地与草业科学、水产科学、热带作物科学等12大学科类建立了一批农业科技基础数据库。二是农业领域野外观测台站的长期定位观测数据,2017年农业农村部启动国家农业科学观测网络建设工作,布局了10个领域的观测实验站456个,附属监测点超过2000个,首次实现我国农业领域多学科协同观测的大联网,持续产生了海量农业科学观测监测数据。三是国家各类科技计划支持的涉农项目和课题研究过程中产生的科学数据集。

农业科学数据具有如下三个特点:一是量大面广且类型多样,包括野外长期定位观测数据、田间科学试验数据、实验室中的实验数据、农业科学调查数据等,即有数值型数据、文本型数据,也有图片、视频、声音等多媒体型数据;二是专业众多且跨度大,涉及作物科学、动物科学、微生物科学、渔业与水产、草地与草业、资源与环境、食品质量安全、农业区划等学科;三是农业科学数据库(集)中的数据量差异比较大,既有记录数超过50万条的数据库(集),如作物品种资源数据库,但也有部分数据库(集)仅有几百条记录。

二、促进农业科学数据开放共享的主要做法

1. 以学科为龙头整合基础数据资源,夯实数据开放共享根基

据统计,我国各类农业研究机构近千家,包括中央级单位、地方科研院所、农业高校系统以及有关涉农科研单位等,每个单位或多或

少都有一些有价值的农业科学数据,但这些数据资源又存在层次不一、交叉重复、质量各异的情况,资源整合是一大难点。在实践中,国家农业科学数据中心以学科为龙头,制定了包括作物科学、动物科学与动物医学、农业微生物科学、草地与草业科学、农业资源与环境科学等12大类核心学科资源整合框架,每个资源大类选择一个资源最多、力量最强的单位作为整合的依托单位,再以多种方式组织该领域其他单位参与资源整合。采用该方法,建立了60余个农业科技基础主体数据库,如作物品种资源数据库、中国饲料数据库等,这些高质量的数据库较好满足了对农业科技基础数据库的应用需求。

2. 以科技计划项目数据汇交为抓手,积极汇聚科研项目数据

分散于国家各类科技计划支持的涉农项目组和课题组手中的科学数据,要实现科学数据开放共享,就必须明确数据拥有者、使用者、管理者的责、权、利,建立科学合理的数据汇交和开放共享机制。国家农业科学数据中心提出了“泛出版”的数据开放共享理念,担起了数据“出版者”和数据中介的角色,数据拥有者按照标准规范对数据进行规范化加工和整理,建立本地科学数据库(集)以及相应的文档,通过科技计划项目科学数据汇交系统开展数据及文档汇交,数据中心对汇交的数据库(集)进行数据审核、赋CSTR唯一标识码、发布共享、以及开展引用评价,数据使用者按照开放共享协议使用数据,并标注数据库(集)的引用。数据中心协同部分用户反馈较好、应用需求较大的数据库(集)与数据拥有者协同,将相应的数据库(集)变成数据论文,推荐在《中国科学数据(中英文网络版)》、《农业大数据学报》上刊载,进一步扩大这些数据库(集)的共享应用。

3. 强化标准规范的制定与应用,促进科学数据的开放共享

要实现农业科学数据资源整合和开放共享,标准规范是基础。根据实践需要,国家农业科学数据中心先后研究制定了《农业科学数据共享管理办法(试行)》《农业科学数据质量控制管理办法》等4项科学数据共享办法,形成了包括《农业科学数据元数据标准》《农业科学数据加工流程规范》《农业科学数据分类分级规范》等15套公共标准、47个农业专业领域标准的农业科学数据开放共享标准体系。

实践表明,标准规范在农业科学数据共享中发挥了重要的作用。面对农业学科繁多,科学数据量大面广,数据类型多样的现状,数据整合显得无从下手。为了解决这个问题,借鉴有关信息组织方法,并结合农业科学数据的特点,及时制定了农业科学数据分类标准,把农业科学数据进行系统梳理,形成12大类、60小类的农业科学数据分类体系,纲举目张,一下子厘清了数据组织的路子。

三、农业科学数据开放共享的成效

经过多年建设实践,大量分散、无序的农业科学数据资源得到有效整合和盘活,充分挖掘和提升了国家原有科技计划投入的效益,农业科学数据开放共享服务体系日渐成熟,农业科学数据资源正在国家农业科技创新和农业农村现代化建设中发挥越来越重要的作用。

农业科学数据资源集聚效应显著。“科技计划项目汇交数据资源”“农业科技基础数据资源”和“农业长期定位观测数据资源”等三类科学数据资源在国家农业科学数据中心实现了有效汇聚,并形成了一批高价值的农业科学数据库(集)。据不完全统计,截止2022年底,国家农业科学数据中心现有数据集(库)17055个,600余个国家科技计划项目数据实现了数据汇交,1921个农业长期定位观测数据集,数据论文49篇。

农业科学数据开放共享效果初现。国家农业科学数据中心形成了覆盖全国的多模式、多渠道科学数据应用与共享服务体系,积极开展“农业科研项目和农业科学家服务”“宏观管理与决策服务”“数据论文出版服务”和“数据管理与分析软件系统和工具服务”等四大类服务。据不完全统计,已为10项国家重点研发计划项目、35项863项目(课题)、22项973项目(课题)、30项国家自然科学基金项目、46项国家科技攻关项目(课题)提供了数据支持。作物遗传资源特性评价鉴定数据库、作物优异资源综合评价数据库、作物品质数据库等数据库有力支撑了农作物核心种质构建、重要新基因发现与有效利用等科学研究,提高了研究效率、加快了研究进程,促进高水平研究成果的产出。

四、未来展望

科学数据开放共享是一项长期的任务。今后将从组织管理、人才队伍、数据汇交、开放共享、数据安全等方面入手,把国家农业科学数据中心建设工作推向一个新的阶段。在资源整合方面,从横向和纵向两个维度不断拓展和深化数据资源的整合范围和深度,逐渐形成国家农业科学数据战略资源长期保障体系,同时实施数据精品化战略,打造一批精品数据库(集)。在数据汇交方面,从标准、技术、工具、系统等入手,帮助数据提供者汇交数据;与重大课题组之间建立良好的互动关系,开展针对性的数据汇交和数据服务工作。在共享服务方面,建立多层次用户服务体系,加强数据服务能力建设,关注用户数据需求,开展用户研究,进一步挖掘数据增值服务,为用户提供深度服务,实现服务品牌化。在运行机制方面,持续推进向“中心实体化、人员专职化、运行规范化、服务常态化”的方向发展。最终把国家农业科学数据中心建成一个“资源丰富、运行稳定、服务高效、数据安全”的全国农业科学数据集散和服务中心,为我国农业科技创新活动提供强有力的数据支撑。

2. 服务层, 由BIPM及各国家计量院 (简称NMIs) 和相关组织实施, 包括基于SI核心表达的《国际计量词汇——基本概念和一般概念》(简称VIM)、测量不确定度表达指南 (简称GUM)、CODATA基本常量、MeP数据库、KCDB数据库、JCTLM数据库、UTC数据库、数字校准证书 (DCC) 元模型XSD等开放数据格式、软件和服务, 按照FAIR原则, 提供数字参考服务。

3. 应用层, 由计量技术机构与全社会的计量伙伴完成, 基于SI核心与服务层的数字化计量服务基础, 丰富计量数字化应用, 如数字校准证书的产业应用、数字测量服务、网络传感器的计量溯源与应用、数字国家基础设施的构建等。

CIPM已与ISO、ISC、CODATA、CIE、IEC、ILAC、IMEKO、OIML等国际组织签署SI数字框架的《联合声明》。该声明为各签署组织提供了一个平台, 作为国际科学和质量基础设施更广泛的数字化转型的一部分, 以适合其特定组织的方式表明其对SI数字框架的开发、实施和推广的支持。

三、计量数据的开放共享

计量数据的开放共享, 首先要实现计量数据和元数据FAIR化的可发现和可访问, 可采用以下方案:

可发现性: 建立计量数据和元数据管理系统, 为每个数据集分配全球唯一、持久的标识符, 如数字对象唯一标识符 (digital object unique identifier, DOI)⁸等, 以确保数据的唯一性并易于搜索。同时, 建立完善的元数据管理系统, 记录数据的关键信息, 包括详细的描述、标签和分类, 以提升数据的搜索和识别性, 也可以让用户全面了解数据的背景和用途。此外, 确保元数据中包含对所描述数据的准确标识符。在核心层的数据管理系统中, 实施可检索的注册和索引机制, 以便数据和元数据能够被检索。这可以通过建立标准化的元数据索引和搜索引擎来实现, 以确保数据可被广泛发现和利用。

可访问性: 采用核心层定义的(元)数据模型进行扩展, 以满足特定领域的需求; 采用标准化通信协议, 如超文本传输协议 (Hypertext Transfer Protocol, HTTP) 或RESTful API, 以标识符进行数据检索。同时, 还需要确保这些协议是免费、开放且可广泛执行的。此外, 协议应该允许在必要时进行身份验证和授权过程, 以确保数据访问的安全性和合法性。同时, 确保即使数据不可用时, 元数据仍然可获取。

四、基于SI数字框架计量数据的开放共享特点

在SI数字框架的SI核心层, “数字SI”工作组及专家组正在准备SI手册的XML版本⁹, 并进行有计划地完善¹⁰。SI数字框架的服务层正在实现可发现、可访问: 创建访问校准和测量能力关键比对数据库的应用程序编程接口 (简称API)¹¹; 确保用于支持关键比较和出版物的数据是FAIR的¹²; 并授予API对机器可读格式的通告T¹³ (BIPM通告T每月发布一次) 数据的访问权限¹⁴。

从SI数字框架和计量数据开放共享各自特点, 以及已经开发的实例, 可以总结出: 计量数据的开放共享在具备机器可操作性的数字化特征的同时, 具有计量的数字可溯源性这一显著特点, 这也是计量数据开放共享的内在要求。

1. 计量数据的计量可追溯性

依据计量溯源性, 计量数据具有天然的计量可追溯性。同理, 在数字世界, 计量数据的数字化也要求建立完整的数字溯源链实现计量可追溯性。对数字计量数据和元数据采用FAIR原则之可发现、可访问, 是实现测量数据计量可追溯性必要条件。

图2显示一张数字校准证书 (DCC) 在SI数字框架中体现的计量可追溯性。由于SI数字框架的三层结构具有明确的层内涵与层边界定义, 因此所建立的追溯链路的各要素间的组织责任明确, 可维护性强。同时, 每一层的要素定义 (特别是SI核心表达层和服务层), 要素与要素之间的溯源关系, 层与层之间的依赖关系, 都隐含了计量追溯性。因此, SI数字框架对计量数据的可追溯性不仅提出了要求, 也提供了指导。

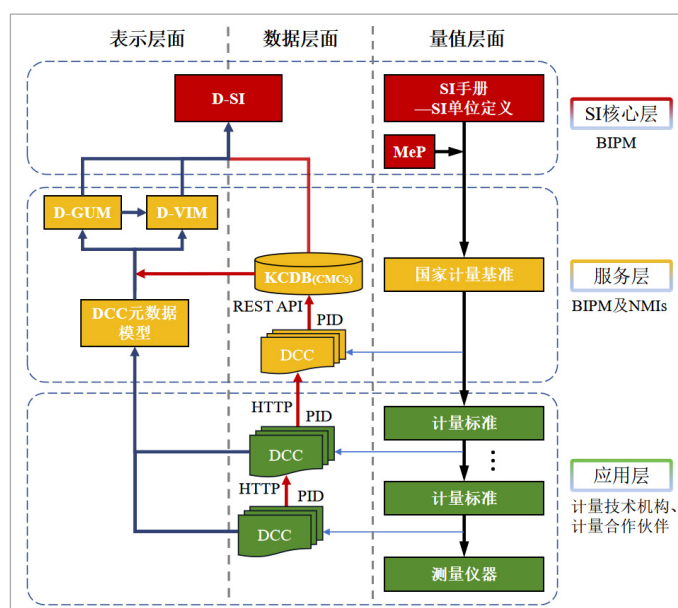


图2 SI数字框架对计量数据可追溯性的保障（以DCC为例）

五、结论

计量数据的机器可读性、可理解性、可操作性，要求实现计量数据、计量报告、计量证书等信息载体的全面数字化，要求其框架统一、格式标准化、操作规范化。计量数据的可追溯性的必要条件是计量数据全面符合 FAIR 的可发现、可访问原则。这两项特点是基于 SI 数字框架的计量数据实现开放共享的显著特点，有助于深入理解基于 SI 数字框架、符合 FAIR 原则计量数据的内涵和特征，有利于实现计量数据的开放共享，推动计量数字化转型。

2. 计量数据的机器可读性、可理解性和可操作性

随着数字化的普及，FAIR原则更多关注于确保数据具备机器可读性，也就是保证计算机在几乎不需要人工干预的情况下能够查找、访问、互操作和重用数据¹⁵。特别是计量数据，在实现互操作性方面，需要明确提供机器可读的测量元数据。这些元数据包括但不限于测量单位、被测物理量类型、测量不确定度形式和测量标准的溯源信息（后二者需在适当情况下，通过校准活动提供）。

SI数字框架的目的不仅仅是以数字形式表示测量单位。它将支持对被测量系统的描述和测量方式，以及与获得的最终测量结果（数据、模型和软件）相关的工作流程。该框架要求提高计量数据的机器可读性水平，从基本的对于计量要素的理解功能到完整的机器可操作知识表示，从而实现数据的机器可操作。

附:简称词表

SI:国际单位制

CIPM:国际计量委员会

BIPM:国际计量局

KCDB:校准和测量能力关键比对数据库

JCTLM:检验医学溯源联合委员会

ISO:国际标准化组织

ISC:国际科学理事会

CODATA:国际数据委员会

CIE:国际照明委员会

IEC:国际电工委员会

ILAC:国际实验室认可组织

IMEKO:国际测量技术联合会

OIML:国际法制计量组织

参考文献:

¹BIPM. Session II of the 108th meeting of the CIPM. [EB/OL]. [2023-09-17]. <https://www.bipm.org/documents/20126/17315032/CIPM2019-II-Decisions-EN.pdf/4d427e8c-22d3-60e6-7eec-1b4b15caece0>.

²CIPM Task Group on the Digital-SI. Draft of the grand vision: transforming the international system of units for a digital world - version 3.4. [EB/OL]. [2023-09-17]. https://www.bipm.org/documents/20126/46590079/WIP+Grand_Vision_v3.4.pdf/aaecfe3-0abf-1aaf-ea05-25bf1fb2819f.

³BIPM. 27th meeting of the CGPM(2022). [EB/OL]. [2023-09-17]. <https://www.bipm.org/en/cgpm-2022/>.

⁴Wilkinson M D, Dumontier M, Aalbersberg I J J, et al. The FAIR Guiding Principles for scientific data management and stewardship[J]. Scientific data, 2016, 3(1): 1-9.

⁵Chalk S J, Coppa D N, Flamenco F, et al. International development of the SI in FAIR digital data[J]. Measurement: wSensors, 2021, 18: 100293.

⁶Mons B, Neylon C, Velterop J, et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud[J]. Information services & use, 2017, 37(1): 49-56.

⁷Stall S, Yarmey L, Cutcher-Gershenfeld J, et al. Make scientific data FAIR[J]. Nature, 2019, 570(7759): 27-29.

⁸McMurry J A, Juty N, Blomberg N, et al. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data[J]. PLoS biology, 2017, 15(6): e2001414.

⁹BIPM. SI Brochure: The International System of Units (SI). [EB/OL]. [2023-09-17]. <https://www.bipm.org/en/publications/si-brochure>.

¹⁰Brown R J C, Janssen J T, Wright L. Why a digital framework for the SI?[J]. Measurement, 2022, 187: 110309.

¹¹BIPM.API KCDB. [EB/OL]. [2023-09-17].<https://www.bipm.org/en/cipm-mra/kcdb-api>.

¹²Lewis A J, Yacoot A, Milton M J T, et al. A digital framework for realising the SI—a proposal for the metre[J]. Metrologia, 2022, 59(4): 044004.

¹³BIPM. Machine-readable data within the context of disseminating the Coordinated Universal Time (UTC). [EB/OL]. [2023-09-17]. <https://www.bipm.org/documents/20126/71876262/DIG-MET-2022-MEYNADIER.pdf/e8e154a7-f527-3320-8f3e-ebd572a9088e>.

¹⁴BIPM. Circular T. [EB/OL]. [2023-09-17]. <https://www.bipm.org/en/time-ftp/circular-t>.

¹⁵Dorst T, Gruber M, Vedurmudi A P, et al. A case study on providing FAIR and metrologically traceable data sets[J]. Acta IMEKO, 2023, 12(1): 1-6.

中国科技期刊开放研究数据政策与实践

孔丽华 习妍

中国科学院计算机网络信息中心

一、科学数据共享在中国的发展

近现代科学的快速发展很大程度上得益于科学的开放式探索。自十七世纪学术期刊创建以来逐渐确立形成了研究结果公开发表与开放交流机制。彼时,作为研究结果证据的数据受限于传播载体的局限,被精简和极度省略,仅仅发布与结果高度相关的数据信息。如今随着信息技术的飞速发展,数据和信息存储、传播等能力迅速提升,翔实的科学数据的公开已具备技术条件。同时,随着数据密集型科研范式的转变,使得它们的公开也变得更加必要和迫切¹。

开放科学数据除了可以提高研究结果的可检验性和公信力以保障科学的自我修正能力之外,还为更多的科学研究提供了丰富的基础资料,产生难以预料的社会和经济价值。因此,科学数据开放共享为国家政府、资助机构、出版机构、公众等利益相关者等广泛关注。

中国开放科学事业发展迅速。自21世纪初以来,中国的政策制定者和资助机构如中国科学院(CAS)和国家自然科学基金委员会(NSFC)不断通过各种国际和全国性倡议,积极推动中国开放获取事业的发展。为了加强开放研究数据分享和管理能力,中国还在全国地区和机构层面上更新其数据政策,围绕开放科学数据开始建设更加全面的学术生态系统²。

2018年,中国国务院办公厅发布《科学数据管理办法》³,针对数据管理的各个方面制定了重要管理办法,其中特别针对数据的出版与传播提出了相关要求“支持科研人员整理发表产权清晰、准确完整、共享价值高的科学数据”,并要求“科学数据使用者应遵守知识产权相关规定,在论文发表、专利申请、专著出版等工作中注明所使用和参考引用的科学数据”。为此,中国科学院、中国农业科学院等机构也根据多年的数据工作经验发布了机构开放数据规则,进一步明确了对学术论文的科学数据汇交要求。如2019年2月中国科学院发布的《中国科学院科学数据管理与开放共享办法(试行)》中明确要求“科研人员应将支持学术论文的科学数据汇交到科学数据管理机构,并适时开放共享,确保科研结论可验证”。同年,科技部、财政部联合公布国家科技资源共享服务平台优化调整名单(20个国家科学数据中心、30个国家生物种质与实验材料资源库),明确推动了对中国科学数据建设、汇聚、整合与开放共享。2022年11月,中国科协、中国科学院决定联合开展论文关联数据汇交工作,发布了《中国科协办

公厅 中国科学院办公厅关于组织开展期刊论文关联数据汇交工作的通知》,鼓励作者在投稿国内科技期刊时,通过“科创中国”科学数据仓储及应用服务平台提供论文关联数据的在线提交、数据评审和开放共享服务。该工作的实施,极大推动我国科技期刊关联数据的出版与共享建设,以及期刊针对对数据的出版实践,进一步促进了各学科间的合作。

此外,中国已经建立了以国家安全为核心的数据安全和个人信息保护的法律框架。并出台了一系列相应的法律法规和标准,涵盖国家网络安全、数据安全、个人隐私保护等方面。

二、中国科技期刊开放研究数据政策与实践

长期以来,发表在科技期刊上的学术论文是科研工作者学术交流的主要形式。科学始于数据,科学数据在科研活动与学术交流中承担越来越重要的角色。通过建立科技期刊数据政策,将学术论文与科学数据进行关联出版,能够有效推动科学数据开放共享,促进数据重用、数据引用和科研评价。

随着中国对数据共享及出版政策制定力度的不断加大,许多学术期刊纷纷制定适合的数据政策。尽管目前尚无具体数据显示中国科技期刊数据共享出版实践的确切数量,但2022年的对中国2019年启动实施的“中国科技期刊卓越行动计划”⁴入选期刊的相关调研^{5,6,7}结果显示,在包括“领军期刊”“重点期刊”“梯队期刊”,以及部分“高起点新刊”在内的302种期刊中,重点调研了相关数据政策的制定及实施情况。其中,已制定相关数据政策的不同类型期刊占比(图1)分别为“领军期刊”81.82%，“重点期刊”62.07%，“梯队期刊”34.67%，“高起点新刊”69.23%。根据中国科协发布的《中国科技期刊发展蓝皮书(2022)》最新统计数据显示,截至2021年底,中国科技期刊总量已达5071种。这302本期刊目前仅占该数据所有期刊的不到1%,但基本代表了中国高水平科技期刊关联科学数据出版当前的业界实践进展。虽然与国际相比,我国科技期刊关联数据出版在行业认可度、框架策略等方面还有待提升,但加快建设学术期刊数据政策,提高数据共享能力,对于中国期刊的影响力至关重要。

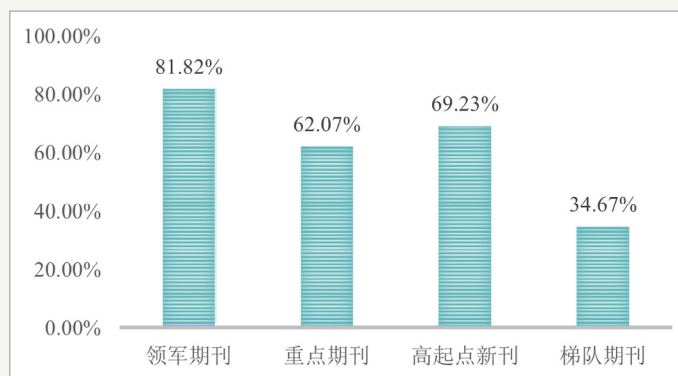


图1 入选“卓越行动计划”期刊关联数据出版政策制定占比情况统计⁷

此外，随着数据出版呈现蓬勃发展态势，数据期刊成为广泛实践与探索数据开放共享的重要手段。作为一种出版形式的探索，科学数据出版先天性地与科学数据的特征紧密关联，在继承媒体出版特色的同时，又为科学数据开放共享提供了新的方法。相比较传统期刊对关联数据的共享出版，这种以数据论文形式的数据出版，通过论文形式组织对数据采集和生产加工过程、数据的组成结构、数据的质量控制和评估方法、数据价值、数据使用方法以及建议等信息，给出了更为深入和丰富的阐述，经过专家同评后正式出版数据论文，确保遵循数据FAIR原则（《FAIR Data Principles》，2016年），达到可发现（Findable）、可访问（Accessible）、可互操作（Interoperable）、可重用（Reusable），更有利于对数据的引用和重用，提供更大的价值。

另有研究^{8,9,10}显示，国际上的纯数据期刊年均数据论文发文量逐年提高，其中尤以Data in Brief和Scientific Data更为显著；而由中国创办的国内两本纯数据期刊，《中国科学数据（中英文网络版）》与《全球变化数据学报》的数据论文发文量处于上升期（图2）。此外，在对科学数据的共享政策方面，包括数据唯一永久标识符、数据可用性声明、数据版权、数据评审、数据引用、数据安全、数据更新处理流程等方面，数据期刊有更加严格的政策，并得以实践¹¹。例如，在《中国科学数据》，我们会对数据论文及所关联的数据实体进行严格检查，努力确保所有已发表论文符合我们所制定的严格质量标准（参见《中国科学数据》投稿须知）。

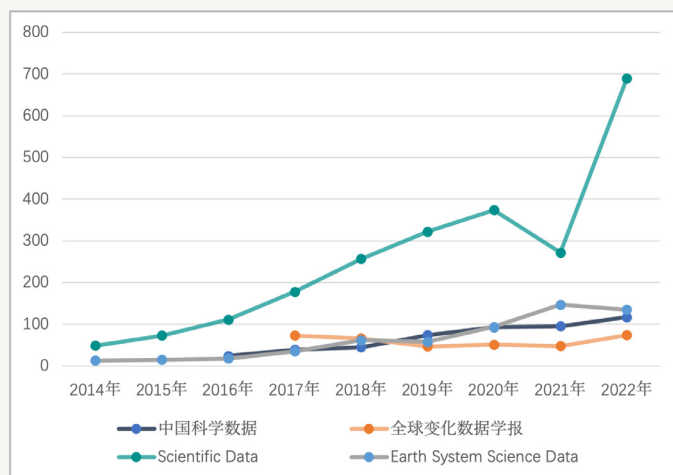


图2 入选“卓越行动计划”期刊关联数据出版政策制定占比情况统计⁷

三、中国科技期刊数据共享基础设施建设

数据的开放需要人工智能、大数据等技术的智能应用，进一步释放数据的价值。科技期刊数据开放和数据期刊发展的支撑基础是开放的数据仓储，科学家或作者将科学数据储存在数据仓储中，并通过对数据进行描述或引用，与研究论文关联集成，从而实现期刊数据开放获取。对标中国政策要求、对接国际服务标准，建设安全可靠的数据仓储平台是推进中国期刊开放数据的最为关键的基础设施¹²。

由中国科学院计算机网络信息中心建设维护的科学数据银行（Science Data Bank, ScienceDB）是一个开放可信的通用型科学数据存储与发布平台，面向全球科研工作者、科研团队、学术期刊及出版商、科研机构及高校等利益相关者，提供科学数据存储、长期保存、出版、共享和获取等服务。平台提供多种数据共享方式与多样的数据许可协议，充分尊重和保障数据所有人权益。ScienceDB致力于出版数据符合主流数据标准或惯例的科学数据，旨在服务全球开放科学数据共享事业，并推动数据共享文化氛围在科研共同体中的培育及良性发展。

为了更好地为我国科技期刊提供服务，科学数据银行在2022年底基于该存储库服务情况制定了一套论文关联数据共享政策模板定制服务¹³，为中国科技期刊制定数据共享政策提供可靠服务。

四、结语

目前,在中国,大量学术出版商和机构都在向开放获取模式转型,也逐步开始关注对科学数据的共享与实践。尽管中国研究人员和研究组织对数据出版的学术性存在一定疑虑,但大多数出版商所表现出的积极性正在逐步打消这种疑虑,包括大力推行论文关联数据的共享政策,强调其在编辑和研究诚信方面的最佳实践标准,积极探索出版数据论文等。未来数年内,随着中国国内期刊市场的不断发展,并在多方的支持下,我们有望看到这些实践在数据共享方面发挥积极影响,在开放科学的出版模式上更上一层楼,从根本上带动大家开放和共享数据的积极性。并通过构建数据开放法律、政策体系,加强数据开放平台建设,不断提升平台易用性,在数据开放的广度、质量方面,进一步提升。以开放促进数据应用,释放数据价值。

参考文献:

- ¹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.
- ²Zhang, L., Downs, R.R., Li, J., Wen, L. and Li, C., 2021. 中国开放研究数据政策和实践综述 (A Review of Open Research Data Policies and Practices in China). Cultural Science Journal, 20 (1), 第3页。
- ³国务院. 科学数据管理办法[EB/OL]. (2018-04-02) [2022-12-28] http://www.gov.cn/home/201804/02/content_5279296.htm. [The State Council, The People's Republic of China. Scientific Data Management Measures [EB/OL]. (2018-04-02) [2022-12-28] http://www.gov.cn/home/201804/02/content_5279296.htm.]
- ⁴中国科学技术协会. 关于下达中国科技期刊卓越行动计划入选项目的通知 [EB/OL]. (2019-11-25) [2022-07-28]. https://www.cast.org.cn/art/2019/11/25/art_458_105664.html.
- ⁵孔丽华, 习妍, 张晓林. 数据出版的趋势、机制与挑战[J]. 中国科学基金, 2019, 33(3): 237-245. DOI: 10.16262/j.cnki.1000-8217.2019.03.005. [KONG L H, XI Y, ZHANG X L. Trends and challenges in research data publishing[J]. Bulletin of National Natural Science Foundation of China, 2019, 33(3): 237-245. DOI: 10.16262/j.cnki.1000-8217.2019.03.005.]
- ⁶孔丽华, 习妍, 姜璐璐. 科技期刊关联数据开放共享及出版政策研究[J]. 中国科技期刊研究, 2022, 33(2): 192-199. DOI: 10.11946/cjstp.202106300526. [KONG L H, XI Y, JIANG L L. Open sharing and publishing policies for research data of scientific journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2022, 33(2): 192-199. DOI: 10.11946/cjstp.202106300526.]
- ⁷孔丽华, 陈枢舒, 习妍, 等. 中国科技期刊卓越行动计划入选期刊数据政策实施现状调研与分析[J/OL]. 中国科学数据, 2023, 8(1). (2023-03-31). DOI: 10.11922/11-6035.ncdc.2023.0002.zh.
- ⁸习妍, 孔丽华, 汪洋, 等. 国内外数据期刊实践与进展研究[J/OL]. 中国科学数据, 2023, 8(1). (2023-03-29). DOI: 10.11922/11-6035.ncdc.2023.0001.zh.
- ⁹王卫军, 李成赞, 郑晓欢, 等. 全球科学数据出版发展态势分析: 基于Web of Science数据库的调研[J]. 中国科学数据, 2021, 6(3). (2021-09-29). DOI: 10.11922/csdata.2021.0019.zh. [WANG W J, LI C Z, ZHENG X H, et al. The analysis of the development trend of global scientific data publishing: research based on Web of Science database[J]. China Scientific Data, 2021, 6(3). (2021-09-29). DOI: 10.11922/csdata.2021.0019.zh.]
- ¹⁰刘凤红, 彭琳. 国际数据期刊的发展现状调查与分析[J]. 中国科技期刊研究, 2019, 30(11): 1129-1134. DOI: 10.11946/cjstp.201904280323. [LIU F H, PENG L. An investigation and analysis of development status of international data journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2019, 30(11): 1129-1134. DOI: 10.11946/cjstp.201904280323.]
- ¹¹张晓林, 沈志宏, 刘峰. 科学数据与文献的互操作[M]// CODATA中国全国委员会编著. 大数据时代的科研活动. 北京: 科学出版社, 2014: 149-158. [ZHANG X L, SHEN Z H, LIU F. Interoperability of Scientific Data and Literature [M]// CODATA, China National Committee. Research activities in the era of big data. Beijing: Science Press, 2014: 149-158.]
- ¹²马瀚青, 关琳琳, 孔丽华, 潘小多, 庞瑜, 何薇, 刘蔚. 数据仓储该如何助推中国科技期刊开放数据?——基于国际科技期刊数据仓储的对比分析. 中国科技期刊研究[J], 2022, 33(4): 470-477 DOI:10.11946/cjstp.202108220669
- ¹³科学数据银行. 科技期刊论文关联数据存缴共享政策定制服务[EB/OL]. (2022-09-03) [2022-12-28]. https://www.scidb.cn/data_policy_tool. [Science Data Bank. Customized Service for Journal Data Policy [EB/OL]. (2022-09-03) [2022-12-28]. https://www.scidb.cn/data_policy_tool.]

开放数据在中国——施普林格·自然的政策、实践和愿景

王重芳 Graham Smith Nick Campbell
施普林格·自然集团

从政策角度来看，施普林格·自然认为，开放出版物背后的数据可以维护研究的完整性、可重复性和透明度。因此，我们支持开放数据的政策和信誉良好的数据存储库。

在中国，我们致力于支持政策制定者履行其职责，确保政策得以实施，并最大限度地提高中国科研的影响力。我们支持所有改善研究数据共享的努力，令人鼓舞的是，国内数据存储库在中国变得越来越普遍，并同时保持着与全球网络的联系。无论位于何处，重要的是所有新数据存储库都应满足公平（可查找、可访问、可互操作和可再利用）共享的全球标准，正如中国科学院的通用数据存储库“科学数据银行”所做的那样。

在实践中，施普林格·自然的愿景是让研究数据作为我们出版物背后的基础证据能够轻松获得，同时赋能我们的作者以最有效的方式共享数据。为了实现这一目标，我们在2023年开展了一些重大研究数据举措，很高兴看到这些举措与本白皮书报告前文中鼓励中国研究人员采用的最佳实践相一致。

特别是今年我们：

启动了研究数据政策的标准化

这将在我们的 3,500 种期刊中嵌入数据可用性声明的要求。此举旨在提高基础数据的透明度，让出版物背后的证据能够被获取，并增强科学记录可信度。作为此次变革的一部分，我们还寻求使我们期刊的作者指南更加直接清楚。

在《自然》旗下期刊中扩大推广 figshare 解决方案

研究表明，将数据共享到数据存储库中的文章平均被引用次数增加 25%¹，这是研究人员共享数据的首要动力。

经过成功的试点后，我们已将figshare数据存储库集成到 35 个《自然》旗下期刊中，其中包括Nature和Nature Communications。这种集成消除了许多作者在论文发表过程中共享数据所面临的挑战，例如必须找到合适的存储库，注册并链接数据等所有繁重的行政工作。

自2022年4月启动该方案以来，figshare已收到7,500份数据提交，相当于手稿提交量的 15%。第一年的数据表明，使用存储库的作者总数增加了 12%。这支持了这样的假设：更易操作的数据解决方案可以对作者的数据共享行为产生明显的影响。我们此计划的下一步旨在提高元数据质量和审稿人对数据的访问。

这一集成服务起源于施普林格·自然与 figshare 长期的合作伙伴关系。自 2014 年以来，双方为支持数据共享推出过许多其他举措，包括 BMC 和 SpringerOpen 期刊的自动上传补充信息、在我们的旗舰数据期刊Scientific Data上增强对作者处理数据的支持，以及 BMC Research Notes 和 BMC Genomic Data 中的内置专家数据检查服务。

除了上述强调的推动有效数据共享行为的解决方案外，施普林格·自然还长期致力于：

确保数据放置在正确的存储库中

我们认为数据应尽可能提交至以学科为基础的、社区认可的存储库。如果作者不知道将数据存放在哪里，我们在网站²上也重点介绍了可以使用的专业和通用数据库的列表。中国科学院的“科学数据银行”与figshare一起包含在通用存储库列表里。中国国家青藏高原科学数据中心则纳入在我们的地球与环境科学数据存储库列表上。

需要说明的是，该列表并不构成施普林格·自然的正式存储库列表，也不具有排他性。其他如 DataCite 的 Repository Finder 和 FAIRsharing registry 中列出的存储库也符合我们的标准。

另一方面，我们对某些特定数据类型（例如大多数组学和晶体学数据）应使用哪个存储库又有强制要求³。例如，新的 DNA 序列、新的 RNA 序列和新的基因组组装数据必须存放到属于国际核苷酸序列合作组织 (INSDC) 中的存储库。考虑到有必要扩大针对这些数据类型的存储库列表，我们在与 INSDC 讨论后，还列出了中国国家生物信息中心所属的基因组序列档案馆 (GSA)，因其正在与 INSDC 合作以实现被后者纳入。

为数据创建和共享提供可引用的、经过同行评审的信用

施普林格·自然有一些开放获取期刊可以发表具有科学价值的数据集的文章类型: Scientific Data 是《自然》旗下一本期刊, 出版较长形式的数据描述类文章。BMC Research Notes, BMC Genomic Data 和相对较新的 Discover 系列所有期刊都能出版较短形式的数据注释类文章。

中国愿景

施普林格·自然自 2016 年 就开始与 figshare 和 Digital Science 合作制作年度开放数据状况报告, 该报告通过全球调研持续提供有关研究人员对开放数据的动机、挑战、看法和行为的详细见解。我们通过这份报告与合作伙伴密切合作, 以了解数据共享并更好地开发解决方案, 这是支持学界一个很好的机会。

在 2022 年的开放数据状况报告中⁴, 我们发表了来自中国科学院计算机网络信息中心的特约文章, 它阐述了中国政策制定者在推动开放数据方面发挥的作用。此外我们发现, 与 2021 年的结果相比, 来自中国的受访者的数量显著增加。2021 年来自中国的调查回复占全球样本的 3%, 而到 2022 年, 该比例上升至 11%, 全球样本总量为 5400。2023 年的调查也收到了同等数量水平的来自中国的反馈。

此次合作之后, 施普林格·自然与中国科学院计算机网络信息中心于 2023 年初又联合举办了网络研讨会 (在线观众超过 22,000 人), 以推广全球和中国的数据共享政策和实践。此次研讨会我们重点介绍了在 2022 年调查中发现的一些中国特有的趋势。

双方对推动中国开放数据进程的良好愿景, 促成了今年共同发布中国开放数据白皮书的深度合作。根据推行最佳实践积累的经验以及过往年度调查的结果, 施普林格·自然完全支持中国科学院计算机网络信息中心在本白皮书第一篇文章中提出的建议。我们希望通过

以下几方面合作进一步支持中国研究人员进行数据共享的实践:

- 推广开放数据政策和建议
- 提供有关资源和实践的培训
- 促成国内数据存储库与国际标准接轨
- 跟踪数据共享的有效性
- 讨论数据共享与学术评价的关系

参考文献:

¹Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. <https://doi.org/10.1371/journal.pone.0230416>

²Data repository guidance: <https://www.springernature.com/gp/authors/research-data-policy/recommended-repositories>

³Mandated data types: <https://www.springernature.com/gp/authors/research-data-policy/repositories-mandates/19540364>

⁴The State of Open Data 2022: https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2022/21276984

作者简介

(按姓氏汉语拼音排序)



Nick Campbell

施普林格·自然学术事务副总裁。Nick于2001年加入当时的自然出版集团，曾任《自然综述：遗传学》编辑、《自然》执行主编，并作为Nature Portfolio总监和《自然》在中国的代
表常驻上海四年。作为《自然》执行主编，Nick领导了该刊至关重要的数字和印刷版的重新发布。在担任现职之前，其最近担任的职务是自然科研执行主编和全球机构合作执行副总裁。Nick的一级学位、博士学位和博士后研究都是遗传学相关领域。他还拥有昆士兰大学的新闻学研究生文凭。



陈润生

中国科学院院士、国际欧亚科学院院士，中国科学院生物物理研究所研究员、博士生导师。主要从事生物信息学研究，是我国最早从事理论生物学、生物信息学以及非编码RNA研究的科研人员之一，曾参加我国第一个完整基因组泉生热袍菌B4基因组序列的组装和基因标识，参加人类基因组“1%项目”和水稻基因组工作草图的研究，构建了国际上在非编码RNA领域有重要影响力的两个数据库，是国内讲授《生物信息学》第一人，曾获中国生物信息学学会(筹)颁发的首届中国生物信息学终身成就奖。



陈昕

博士，中国科学院科学数据总中心副主任，中国科学院计算机网络信息中心高级工程师、大数据部科学数据体系架构实验室主任，国际研究数据联盟Fair Digital Object Fabric组联合主席，长期从事科学数据管理、分析、共享和领域应用工作，研究兴趣包括科学数据治理框架与互操作技术等。



顾昕

现任职国家知识产权局知识产权发展研究中心首席研究员。兼任中央财经大学知识产权研究中心研究员，中国科协财政项目评审专家，中国地理学会“科创中国”优质地理产品
生境保护与可持续发展专业科技服务团顾问组成员，北京市法学会科技法研究会常务理事、副秘书长，中国技术经济学会知识产权专委会理事。



胡晓彦

中国科学院国家空间科学中心高级工程师，硕士生导师，中国科学院青年创新促进会成员。主要从事科学数据治理与空间科学大数据智能应用研究，组织团队在科学数据治理概念建模、空间科学数据互操作、基于机器学习
的知识挖掘等方向开展关键技术突破。主持或参与国家自然科学基金、国家重点研发计划、国家重大科技基础设施、中国科学院战略性先导科技专项、中国科学院信息化专项等多个科研项目。



姜璐璐

工程师，中国科学院计算机网络信息中心数据出版实验室副主任，长期从事科学数据管理与科学数据出版工作，现工作于中国科学数据银行 (Science Data Bank, 简称
ScienceDB)，负责ScienceDB产品运营与合作。在科学数据管理、共享与出版实践上具有长期实践经验。



阚侃

硕士, 高级工程师, 国家一级注册计量师, 现就职于华南国家计量测试中心/广东省计量科学研究院科研部, 主要从事计量数字化基础设施及关键技术、智能控制系统数字化测量等方面的研究。



孔丽华

博士, 高级工程师, 中国科学院计算机网络信息中心信息化发展战略与评估中心副主任,《中国科学数据(中英文网络版)》及《数据与计算发展前沿》编辑部主任, 国家新闻出版署医学期刊知识挖掘与服务重点实验室专家委员会

委员, 中国科学技术期刊编辑学会国际交流与合作工作委员会委员。主要从事科学数据管理政策、数据开放共享政策与实践, 以及数据出版等方面的研究。近年来, 主持及参与完成了多个重要项目及研究课题。作为负责人, 带领团队获得第五届中国出版政府奖“先进出版单位奖”。



李国庆

中国科学院空天信息创新研究院研究员, 国家对地观测科学数据中心主任, 科技部综合对地观测数据共享平台负责人, 科技部国际灾害数据应急援助机制(CDDR)负责人。研究领域为高性能地学计算、网络化数据工程

和灾害数据工程, 发表SCI/EI等论文150多篇, 专著多部, 负责和主持过五十多项国家级重大研究项目, 获2021年度地球观测组织(GEO)卓越个人奖。



李宗闻

工程师, 就职于中国科学院计算机网络信息中心。目前从事科学数据出版的研究工作, 包括科学数据仓储平台建设、学术期刊论文关联数据汇交共享机制等; 负责科学数据银行 ScienceDB 的宣传推广工作, 以及期刊、机构

等数据社区需求分析和建设推广工作。参与了中国科协“科技论文关联数据仓储及应用服务平台”“期刊论文关联数据汇交工作支撑服务”等课题研究。



刘宁

中国科学院计算机网络信息中心高级工程师, 主要从事科学数据治理、元数据互操作与集成应用方面的理论与系统建设, 主要参与中国科学院科学数据库工程系列标准、

中国科学院科学数据中心体系标准体系的设计与标准研制, 主笔相关标准数十项, 承担中国科学院科学数据云门户系统、中国科学院项目科学数据汇交管理平台等科学数据治理与集成服务系统建设。



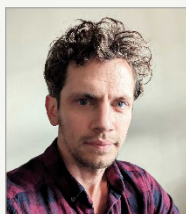
刘洁

中国科学技术大学公共事务学院2021级法学硕士研究生, 研究方向为知识产权。主要从事数据知识产权相关研究工作, 参与数据知识产权保护规则研究报告撰写, 走访学会调研, 参与“科学数据治理法律服务项目”等课题研究。



马俊才

博士, 研究员, 现任国家微生物科学数据中心主任, 中国科学院微生物研究所微生物资源与大数据中心主任, 世界菌种保藏联合会 (WFCC) 执委、世界微生物数据中心 (WDCM) 主任、中国生物工程学会生物技术与生物产业信息中心主任、世界微生物菌种保藏联合会执委、亚洲研究资源网络数据管理工作组主席、国际生命条形码项目数据镜像工作组共同主席。



Graham Smith

施普林格·自然开放数据项目经理。他的主要职责是针对全公司的出版活动, 开发数据共享工具, 推动相关合作, 制定开放数据相关的政策和举措。他拥有地球物理学背景, 并协调组织了Nature、BMC和Springer旗下期刊以及伦敦自然历史博物馆的数据管理活动。



汤惟玮

助理研究员, 国家空间科学数据中心办公室主任。主要从事国内外空间科学领域数据政策研究及分析、各国航天机构及领域国际组织战略规划跟踪分析与开放科学实践研究等。参与国家自然科学基金、中国科学院信息化专项、中国科学院院重点部署项目、中国工程院战略研究与咨询重点项目、中国科学院战略性先导科技专项 (A类) 等多个科研项目。



佟继周

中国科学院国家空间科学中心研究员, 硕士生导师, 空间大数据技术研究室副主任, 国家空间科学数据中心副主任。主要从事空间科学数据工程与大数据应用技术研究。主持或参加空间科学战略性先导专项、国家重点研发计划、子午工程、国家科学数据共享工程、院创新性方向课题等多个科研项目。曾获军队科技进步一等奖, 被授予“中国科学院加载人航天工程优秀工作者”荣誉称号。



王重芳

施普林格·自然集团大中华区开放科学总监。主要负责施普林格·自然开放科学与开放获取相关政策交流及战略合作, 与相关政府及学术机构保持良好互动。在此之前, 她在学术出版领域的不同出版机构拥有 15 年的丰富经验, 从而对出版格局新发展有着深刻的理解。她的专长是开放获取、开放科学、期刊出版及战略合作。



熊行创

博士, 硕士研究生导师, 中国计量科学研究院研究员, 享受国务院“政府特殊津贴”专家, 现任中国计量科学研究院数据中心副主任、国家计量科学数据中心常务副主任。主要从事计量数字化基础设施及关键技术、标准参考数据、科学数据FAIR化、数据质量测量等方面的研究。



胥鑫

中国科学院国家空间科学中心工程师。主要从事数据开放出版研究,在开放共享关键技术、开放出版平台建设及开放科学与开放数据等方面开展工作。参与国家自然科学基金、国家重点研发计划、中国科学院科学传播专

项、中国科学院信息化专项等多个科研项目。



许琦

中国科学院国家空间科学中心工程师。主要研究方向为科学数据标准化与空间科学数据管理,在科学数据标准体系、空间科学数据汇交管理、科学数据仓储理论架构、领域开放科学实践;曾参与WDS元数据收割工作组、全球

开放科学云(GOSC)等相关工作小组;主持或参与多项科学数据领域国家标准、团体标准研制;主持或参与国家自然科学基金、国家重点研发计划、中国科学院信息化专项、国家科技基础条件平台中心委托任务多个科研项目。



张泽钰

工程师,现就职于中国科学院计算机网络信息中心大数据部,本硕攻读法学专业,硕士研究方向为知识产权法,持有国家法律职业资格证书、工商管理方向经济师证书。在此之前,有两年海外工作经历,任职于世界五百强

企业,主要从事法律咨询、合同管理等工作。长期从事知识产权法、网络法领域的研究工作,目前主要研究方向为科研数据权益管理和数据合规。



周国民

理学博士,研究员,中国农业科学院农田灌溉所所长、党委副书记,国家农业科学数据中心主任,中国农业科学院数字化南繁技术创新团队首席科学家,兼任中国农学会计算机农业应用分会副理事长、中国仿真学会农业建

模与仿真专业委员会主任。长期从事农业科学数据共享技术与系统等研究工作,取得较为丰富的科技成果,曾获得各级科技成果奖励12项。



周园春

博士,研究员,博士生导师,现任中国科学院计算机网络信息中心副主任、科技委员会主任,中国科学院科学数据总中心主任,大数据分析系统国家工程研究中心副主任,中国信息协会科学数据专业委员会主任,中国档案学会科技档案与科学数据管理学术委员会副主任。研究

领域为科学数据与数据智能等,在IJCAI、AAAI、IEEE TKDE、ACM TKDD、Nucleic Acids Research等国际著名期刊和会议,发布国家标准5项,出版专著3本,授权专利40多项。



邹自明

中国科学院国家空间科学中心研究员,博士生导师,国家空间科学数据中心主任,中国科学院特聘研究员。主要研究领域为空间科学信息学,在空间科学数据处理,宇宙空间信息的组织、检索与互操作,空间信息系统工程,

大数据与人工智能技术在领域知识发现中的应用等方面开展研究。曾获军队科技进步一等奖两项,中国科学院载人航天工程重要贡献奖和中国科学院杰出科技成就奖等奖项。



中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



科学数据银行
Science Data Bank

中国开放数据白皮书 2023



SPRINGER NATURE

