



人口神经科学数据共享: 政策生态、基础设施、实践和挑战

姜璐璐¹, 高鹏², 周园春^{1,3*}

1. 中国科学院计算机网络信息中心, 北京 100083;

2. 北京师范大学认知神经科学与学习国家重点实验室, 北京 100875;

3. 中国科学院大学杭州高等研究院, 杭州 310024

* 联系人, E-mail: zyc@cnic.cn

2024-02-04 收稿, 2024-05-28 修回, 2024-07-26 接受, 2024-07-30 网络版发表

摘要 随着开放科学时代的到来, 全球的数据共享政策日趋完善, 科学数据共享逐步成为学术交流的必选项. 科研模式及学术交流模式正发生着深刻的变革, 如何让科学数据开放共享可持续发展, 依然面临着来自文化、政策与实践路径上的现实阻力和挑战. 本文梳理了科学数据开放共享可持续发展的必要组成及彼此间的作用关系, 分析了利益相关方在数据共享中的诉求与需要付诸的投入. 接着, 从各方主体的科学数据共享政策情况、科学数据共享的标准及原则、数据存储库的标准规范指南、人口神经科学领域数据存储平台建设实践情况等方面, 结合我国“科学数据银行: 中国人彩巢计划数据社区”进行具体介绍与特征分析. 人口神经科学领域的快速发展在很大程度上得益于科学数据共享, 但依然面临着来自隐私安全与伦理, 数据共享激励机制缺失, 科研群体科学数据管理素质不足等突出问题和挑战. 其中, 伴随着新型科学技术的发展, 隐私安全正被不断地完善和保护. 针对科学数据共享的发展现状及挑战, 本文基于在数据共享的政策、建立激励机制、基础设施建设、跨学科融合发展等方面的研究提出人口神经科学在科学数据共享工作上的未来发展趋势与建议.

关键词 科学数据共享, 数据政策, 数据标准规范, 科学数据银行, 数据社区

大数据战略的落实将极大地促进实现新一波的未来人口和人才红利, 实现人口高质量、科学化发展, 是国家富强和民族复兴的重大需求. 发展人口神经科学^[1]以其在纵向追踪、人口代表性、大数据应用、人工智能与创新转化等方面的独特优势成为脑科学领域前沿, 使神经影像向教育学、心理学和医学的转化应用成为现实^[2]. 发展人口神经科学的学科发展和创新应用转化将通过建立脑影像、基因和环境等大规模纵向研究队列大数据, 通过多学科交叉融合揭示基因和环境如何塑造人脑发展的一般规律, 为实现人口素质提升提供科学数据共享基石^[3,4]. 实现这一目标, 离不开跨学科

的海量数据(除脑影像数据外, 还包括环境的、人口的、地理的等多学科数据)支撑以及开放的数据共享环境.

学科的自身发展需要数据共享. 数据共享在提升研究可复现性、较低科研成本等方面具有诸多益处. 首先, 数据共享能够有效提升研究的可复现性, 确保科学自纠正能力和公信力. 缺乏可复现性是心理学研究面临的一个问题^[5]. 2015年引发的心理学研究信任危机^[6-8]让科学家们开始关注研究的可复现性问题, 心理学、神经科学及相关领域研究也迎来了关于增强研究可复现性的“复兴运动”^[9,10]. Goodman等人^[11]在2016年

引用格式: 姜璐璐, 高鹏, 周园春. 人口神经科学数据共享: 政策生态、基础设施、实践和挑战. 科学通报, 2024, 69: 3560–3577

Jiang L L, Gao P, Zhou Y C. Population neuroscience data sharing: Policy ecology, infrastructure, practices and challenges (in Chinese). Chin Sci Bull, 2024, 69: 3560–3577, doi: [10.1360/TB-2024-0147](https://doi.org/10.1360/TB-2024-0147)

的分析认为, 研究的可复现性包括了方法可复现性(methods reproducibility), 结果可复现性(results reproducibility), 以及推断可复现性(inferential reproducibility)。科学数据, 乃至分析软件及代码, 都是实现可复现性的必要支撑。开放数据使得研究的原始观点可被他人审查验证, 有助提升研究的可靠性^[5]。此外, 科研数据共享能够有效降低科研成本, 激发数据再利用价值, 推动科研共同体的合作共赢。研究者可利用他人公开的脑影像数据测试他们的假设, 而不是从零开始收集数据^[12,13], 节省资金^[14]的同时, 加快了认知神经科学的研究进程。以“国际千人功能连接组计划(1000 Functional Connectomes Project, FCP)^[15]”为例, 2008年, 来自全球35个脑影像实验室联合发起此项计划。该计划通过美国网络信息化平台(The Neuroimaging Tools and Resources Collaboratory, NITRC)共享了1414名志愿者的脑影像数据, 引起了神经科学领域的广泛关注并推进了人脑功能科学探索进程, 为人脑功能大数据研究提供了可能。2009年, 人类脑连接组计划(Human Connectome Project, HCP: 2009~2014)^[16]启动, 旨在通过高分辨率神经影像技术绘制人脑结构和功能连接图谱, 通过数据共享提升了科研的透明度和可合作性, 推动了脑连接组学研究, 促进了对人脑功能和疾病机制的深入理解, 为神经科学领域的基础研究和临床应用提供了宝贵的资源。FCP团队核心成员发起“国际神经影像数据共享倡议(International Neuroimaging Data-sharing Initiative, INDI)”, 并成立战略规划委员会, 在国际上引起广泛关注, 促进了人脑神经影像研究的开放化。基于HCP的经验和技术积累, 美国启动了“青少年脑与认知发育计划(Adolescent Brain Cognitive Development, ABCD)”国家健康战略计划, 该计划将对全美1万名9~10岁儿童进行10年随访, 聚焦药物滥用的脑发育机制^[17]; 此外, 英国建立了“生物银行(UK-Bio-bank)^[18]”, 覆盖10万中老年人的活体脑成像数据。实践表明, 数据共享已成为大型活体脑成像计划的重要计划目标, 并有效发挥了数据重用价值。更有学者提出数据共享是研究人员的道德义务^[19]。

以脑影像研究领域为典型代表的科研共同体, 自驱地开展了开放数据共享实践, 并助力开放科学数据走入更广泛的科研群体、资金资助方及政府、公众视野。2021年联合国教科文组织推出《开放科学建议书》(<https://www.unesco.org/en/open-science/about>), 开放科学进入全球共识新阶段, 开放科学数据作为其中

的重要组成部分也受到了广泛关注。在政府层面, 美国白宫科技政策办公室(The White House Office of Science and Technology Policy, OSTP)将2023年定为开放科学年(<https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/>), 推进国家开放科学政策, 开放数据是其中的重要组成部分; 2023年1月1日美国NIH的数据政策生效, 要求公开共享数据^[20]; 2017年起, 欧盟在地平线2020计划中明确要求资助项目数据存储在合适的存储库; 英国、澳大利亚^[21]、金砖国家^[22]等政府及政府组织也在科学数据管理、科学数据开放共享上开展诸多实践。中国于2018年颁布《科学数据管理办法》(https://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm), 提出“开放为常态, 不开放为例外”的公共资金资助的科学数据管理原则; 2019年首次提出“将数据作为生产要素”, 2023年中国政府组建了国家数据局, 统筹数据资源整合共享和开发利用(http://www.news.cn/politics/2023lh/2023-03/07/c_1129419141.htm), 并印发《“数据要素×”三年行动计划(2024—2026年)》, 在“数据要素×科技创新”中提出“推动科学数据有序开放共享”。

在学术交流层面, 国际科学、技术与医学出版商协会(International Association of Scientific, Technical and Medical Publishers, STM)将2020年定义为“科学数据年(Research Data Year)”(<https://www.stm-assoc.org/standards-technology/2020-stm-research-data-year/>)。国际主要出版商, 包括Springer Nature、Elsevier、Wiley、Taylor & Francis等在内的出版商先后出台数据政策呼吁论文出版时共享相关支撑数据; 2023年3月, 美国《科学》杂志发表题为“强化科学记录”的社论, 表示为履行“研究可复现性”和“FAIR数据原则(Findable, Accessible, Interoperable & Reusable, FAIR Principles)”的承诺, 《科学》将“要求发表论文中作为结果的所有基础数据公开并即时可用”^[23]。

在开放科学运动的推动下, 传统科研模式及学术交流模式正发生着深刻的变革。在这种变革下, 如何让科学数据开放共享可持续发展, 以实现增强公众信任、节约科研成本、释放数据价值及促进科学研究这些目标, 依然面临很多现实阻力和挑战, 包括有效实施路径, 共享意愿与文化, 政策环境等多方面因素。

1 科学数据开放共享可持续发展的必要组成与生态构成

实现可持续的数据共享,根本上需要满足各利益相关方的诉求.科研共同体需要高质量的科学数据供给,服务自身科研工作需要,复现他人研究成果,发掘新的合作机会等.数据贡献者也需从共享行为中获益,如资金支持、成果认证、职业晋升与学术声誉等.出版商/期刊需要通过推动作者共享论文支撑数据来提升发表成果透明度,以规避学术不端、扩大论文影响力等实现期刊声誉建设.政府、资金资助方、机构等政策制定方则需从开放数据中实现科研数据资产管理、投入产出效益最大化的诉求,以惠及社会与公众.

基础设施、服务平台、共享标准、激励机制与配套政策是数据共享可持续发展的必要组成部分,这些组成部分的关系如图1示.数据共享基础设施保障是第一层基础.数据共享基础设施包括有存储、数据长期保存机制与网络等,是满足数据文件可长期安全存储、有效管理及可远程传输的必要保证,是让数据共享成为可能的基础性保障.随着网络、云计算等技术的快速发展,数据共享基础设施能力已不再是近年来

数据共享工作的主要阻碍因素.基于高可用的数据共享基础设施保障,数据共享服务平台是第二层基础.通过对数据文件的描述、数据管理及对外服务等方式,数据共享服务平台使得共享数据可被索引发现及访问.配套标准规范是提升共享数据可被理解、被使用以及被重用的重要保障,对发挥科学数据的利用价值起重要作用;同时,数据共享的标准规范有利于提升科研群体对数据共享的认知与认同,提升科学数据管理贯穿科研全生命周期的实施,也让数据资产管理、数据价值评价等成为可能.在数据可利用的基础上,赋予数据共享的激励机制让数据共享行为被认可,对于构建良性的数据共享文化、培养可持续的数据共享科研习惯十分重要;来自政府、资金资助方、机构、期刊等的政策才能让科学数据共享成为科研过程中的必要环节.各部分间自下而上起到支撑保障作用,反之,上层的发展对下层起到了促进和提升作用.

一个良性可持续发展的科学数据共享生态离不开各利益相关方的共同参与和贡献.底层基础的保障建设(图1第1~2层)须政府、资金资助方及机构等进行长期的资金投入,数据服务平台的数据治理能力、数据共享的标准规范需要科研共同体发挥关键作用.而配

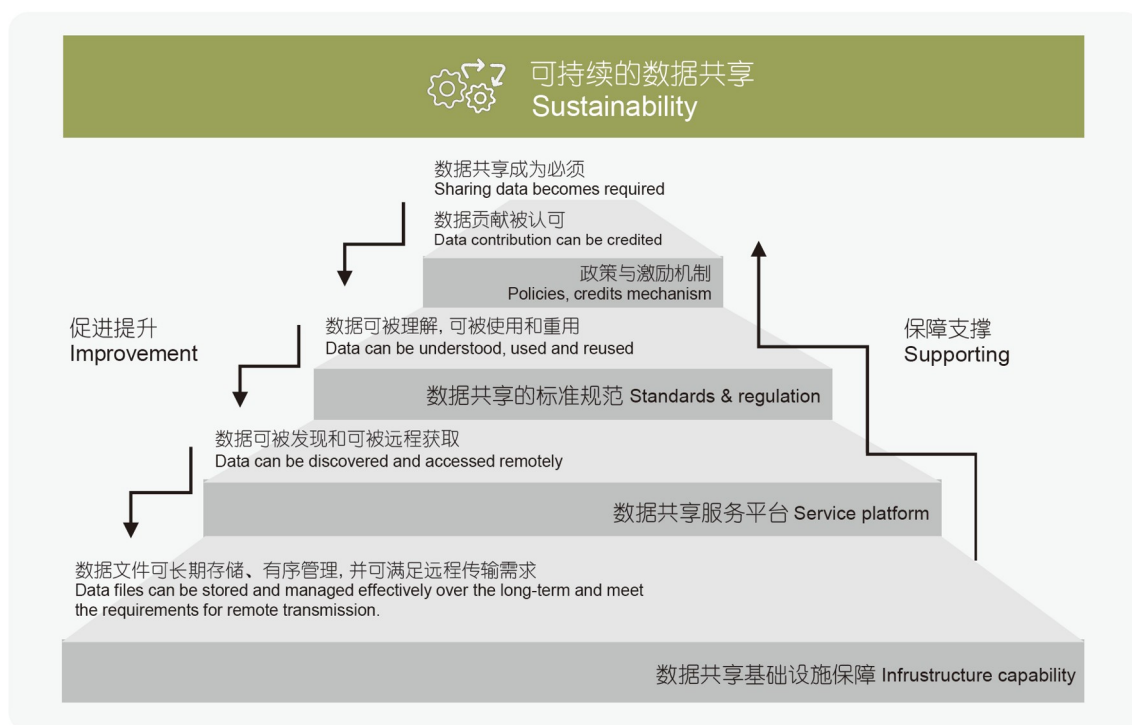


图 1 (网络版彩色)数据共享可持续发展的必要组成

Figure 1 (Color online) Necessary components of data sharing for sustainable development

套政策与激励机制的建设形成, 需要包括出版商/期刊、机构、资金资助方、政府等在内的政策制定方联合科研共同体一起构建, 以形成制度牵引和各方共识。数据贡献者在数据共享中付出的投入最为关键, 需要在科研活动中开展符合标准规范的数据管理工作, 因此, 足够的激励与健全的政策保障机制是促进共享行为最主要的动力。

在开放科学运动的浪潮下, 科研共同体自下而上地推动着数据共享实践, 系统地研究和梳理数据共享中各必要组成部分的发展现状及问题挑战, 对推动数据共享的良性可持续发展, 促进发展人口神经科学及创新应用十分必要。随着开放科学已成为全球共识, 以顶层设计为牵引的数据共享生态建设模式逐步形成并发挥起重要作用。

2 科学数据开放共享的政策现状

2.1 政府与资助方的政策情况

2003年, 欧盟出台《欧洲议会和理事会关于公共部门信息再利用指令》(PSI指令), 并于2013、2019年完成2次动态修订。2013年欧盟发布《关于Horizon 2020计划中的“科学出版物和研究数据开放获取指南》(https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm), 并启动开放研究数据试点(Open Research Data Pilot), 从2017年起覆盖所有领域, 要求项目将研究数据存储在适当的存储库中, 旨在提供对Horizon 2020项目产生的研究数据的访问和再利用^[24]。2021年6月, 科学欧洲(Science Europe, SE)发布《可持续科研数据实用指南》(Practical Guide to Sustainable Research Data, <https://zenodo.org/records/4769703>), 为数据的长期保存和可访问性提出指导, 涵盖资助者、执行者和数据基础设施的三个互补成熟度矩阵。2023年底, 欧盟发布连续第9年发布“开放数据成熟度报告”(2023 Open Data Maturity Report, <https://data.europa.eu/doi/10.2830/384422>), 从开放数据政策、开放数据影响力、开放数据门户、开放数据质量四个维度的12个指标针对35个欧洲国家的开放数据成熟度情况展开评估^[25], 欧盟各成员国的平均政策成熟度已达89%。

英国政府对科学数据管理及共享十分重视。早在2003年英国成立数据管理中心(DCC), 促使英国国家数

据管理机构成立及科学数据开放共享的实践。2012年英国皇家学会发布《科学作为一项开放的事业》^[26], 分析了开放科学数据的重要意义并倡议付诸实践。2016年英国政府发布《开放科学数据协议》, 2022年颁布了《数据共享治理框架》(Data Sharing Governance Framework, <https://www.gov.uk/government/publications/data-sharing-governance-framework>), 各大资助机构频繁修订数据政策, 仅在2022半年内就发布了近十项数据政策^[27]。

2022年8月25日, 美国白宫科技政策办公室(OSTP)发布“确保免费、立即和公平地获取联邦资助的研究”备忘录(下简称“2022 OSTP Memo”, <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>), 要求所有由联邦资助的科研人员都需要考虑公开研究数据和其他研究产出成果; 各机构制定新计划, 确保到2025年同行评审出版物和联邦资助研究产生的相关数据在出版之日起立即免费提供。作为2022 OSTP Memo的落实执行, 美国国家科学基金会National Science Foundation, NSF, https://nsf.gov/resources.nsf.gov/2023-06/NSF23104.pdf?VersionId=cSTD31SSPUEkM_Vm25HSIlgZBDeiPvzdQ)^[28]、美国国立卫生研究院(National Institute of Health, NIH, <https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-implements-data-management-sharing-policy>)纷纷出台政策推进工作落实。此前, NIH下属美国国立精神健康研究所(National Institute of Mental Health, NIMH, <https://grants.nih.gov/grants/guide/notice-files/NOT-MH-23-100.html>)已发布数据共享政策通知要求2023年1月25日之后的涉及人类受试者的申请都应包括资源共享计划。

中国政府于2018年颁布《科学数据管理办法》; 2019年6月, 科学技术部、财政部部署了20个国家科学数据中心, 覆盖农林、气象、空间、天文、海洋、人口健康、微生物、地球系统等多个领域^[29]; 2019年科技部制订了《科技计划项目科学数据汇交工作方案(试行)》(国科办基(2019)104号), 从项目管理层面建立起科学数据质量控制体系。

2.2 期刊与出版商的政策情况

2020年被STM定义为“科研数据年”, 提出了“共享-链接-引用”的目标, 以及出版商、期刊、数据仓储平台等参与方的最佳实践原则, 全球有21个国际出版商,

13,064种期刊参与其中^[30]。2023年,STM、DataCite以及Crossref发表联合声明,提出了“研究数据共享的清晰路径”,覆盖研究人员、数据仓储平台、出版商、资助方、研究机构以及政策制定机构(<https://www.stm-assoc.org/research-data-program/bestpractices/>)。Springer Nature、Elsevier、Wiley、Taylor & Francis等国际出版商制定了自己的数据政策,有些采用分层级框架,有些实施统一的数据政策(表1)。2023年3月, Springer Nature宣布全面改革其数据政策,弃用此前的4级政策框架转为统一政策,以更有力的举措推动所有出版形式的数据共享(<https://group.springernature.com/gp/group/media/press-releases/springer-nature-makes-data-sharing-easier-with-single-policy/25211176>)。此外,出版商们通过数据论文、开放数据徽章等方式使得数据共享得到认可。

2021年6月及2022年6月中国科学院先后对其主管的400余种科技期刊提出论文关联数据的管理工作要求,鼓励科技期刊关注和探索开展科学数据共享工作。2022年11月,中国科学技术协会办公厅联合中国科学院办公厅发布《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》,鼓励期刊建立数据政策。从期刊核验工作的角度,国家新闻出版总署连续三年(2021~2023)将期刊“是否开展数据管理”纳入出版核验项,越来越多的国内期刊推出了自己的数据政策^[31~33]。

在提升研究透明度上,神经科学领域的期刊提出了具体要求。一是要求研究人员提供完整的原始数据和使用算法,以便研究结果可被复现和验证。二是鼓励研究者使用标准化的数据格式和公共数据库,以更容易地实现数据共享的FAIR原则。一些期刊还开始要求研究者在提交论文时,提供数据处理的详细方法,包括使用的统计模型和软件版本。例如,在进行大脑成像研究时,研究者需要明确指出使用的成像技术、数据处理软件和参数设置等,确保他人可准确复制实验。

2.3 国际组织和国际计划推出的数据共享倡议

科研共同体作为科学数据共享的第一参与方和获益方,通过各种国际学术组织和国际计划倡导科学数据开放共享,为政府、政府间国际组织的数据政策制定工作提供重要参考,并起到重要的实践推动作用。

国际科技数据委员会(Committee on Data for

Science and Technology, CODATA)于2019年发布《研究数据北京宣言》(The Beijing Declaration on Research Data, <https://codata.org/the-beijing-declaration-on-research-data/>),涉及数据管理能力和数据政策体系建设的必要性,公共资金资助的科研数据应尽可能在全球共享重用及宣言落地建议举措等十条原则。

人脑图谱组织(Organization for Human Brain Mapping, OHBM)理事会于2014年发布了“神经影像研究与数据完整性的声明(Statement on Neuroimaging Research and Data Integrity)”,并成立“数据分析与共享最佳实践委员会(Committees on Best Practices in Data Analysis and Sharing, COBIDAS)”,致力于提供可灵活选择的数据管理与共享实践方案,并随着领域的发展而不断进化^[34]。此外,“荟萃分析提升神经影像遗传学(Enhancing Neuro Imaging Genetics through Meta Analysis, EMIGMA)”联合体集成了世界各地的大规模脑成像和遗传数据,允许跨国界的影像研究并鼓励数据共享和科研成果的开放获取,以促进科学知识的普及和进一步的研究^[35]。在具体疾病研究上,“帕金森病进展标记物联合体(Parkinson's Progression Markers Initiative, PPMI)”是一个典型国际性的研究项目。通过各国科研人员的合作和共享,PPMI提供了一个全球性平台,用于识别和验证帕金森病进展的生物标志物,并寻找治疗帕金森病的新方法^[36]。这种国际性合作对于提高研究的广度和深度、确保数据的多样性和质量,以及促进全球性医疗进步具有不可估量的价值。

政府、资金资助方、学术出版商与期刊、国际组织等在政策体系建设中从不同的视角出发,总体上彼此互相补充及调和,比如政府政策偏向顶层设计的框架搭建和原则制定等,资金资助方、机构在顶层框架和原则指导下侧重项目周期管理与数据资产管理,出版商着眼推动学术成果交流,国际组织侧重推进国际合作与共赢,平衡不同国家政策间的差异与统一等。

但近年来的实践表明:仅靠鼓励性的数据共享政策,往往得不到科研人员的有效响应。Springer Nature的一项统计显示,科研人员虽然支持数据开放共享,但只有不到40%的作者在发表论文时主动公开他们的研究数据。2023年,从政府到出版商都相继出台了更为严格的,甚至强制性的数据共享政策。这不仅表现了各方在推动数据共享上的决心,也反映出多年来仅靠鼓励性的数据共享政策,并未发动更多的科研人员参与到开放数据的科学实践中。

表 1 国际主要出版商的数据共享政策一览表

Table 1 List of data sharing policies of major international publishers

出版商	数据共享政策情况一览	数据存储政策	数据期刊
Springer Nature	单一政策, 要点包括: (1) 旗下所有期刊要求作者必须提交数据可用性声明; (2) 强烈鼓励论文出版时公开相关支撑数据, 科研社区支持的数据类型要求强制公开; (3) 审稿人有权要求访问研究的基础数据及代码; (4) 不宜开放共享的数据须在数据可用性声明里说明获取条件.	向作者提供推荐数据仓储库列表; 指定部分数据类型强制使用的 仓储库列表	<i>Scientific Data</i> <i>BMC Genomic Data</i>
Elsevier	单一政策, 要点包括: (1) 鼓励和支持在适当的情况下尽早共享数据; (2) 鼓励和支持正确的数据引用实践; (3) 建立数据审查实践; (4) 支持将科学数据以单独的、经同行评议的成果发布等.	向作者提供推荐数据仓储库列表	<i>Data in Brief</i>
Wiley	分级政策(4级): (1) 鼓励数据共享; (2) 期望数据共享; (3) 强制要求数据共享; (4) 强制要求数据共享以及数据同行评议.	未推荐具体的数据仓储库	<i>Geoscience Data Journal</i>
Taylor & Francis	分级政策(5级): (1) 基本数据共享政策; (2) 据合理请求共享数据; (3) 公共可获取; (4) 公开数据; (5) 公开和FAIR共享数据.	向作者提供推荐数据仓储库列表	暂无

3 科学数据的相关标准及指南现状

3.1 科学数据管理的相关标准及指南情况

科学数据的价值取决于利用^[37]. 科学数据的相关标准及指南的研究, 对于充分发挥科学数据的利用价值, 确保提升研究可复现性和数据可重用性等具有重要意义; 同时, 对支撑政策落实, 对指导科研人员、管理者及服务平台建设方等提升科学数据管理能力和数据服务能力都有重要作用.

在科学数据管理方面, Wilkinson等人^[38]于2016年提出了关于数据共享的FAIR原则, 并快速获得全球科学共同体和各国政府的认可和支持. 在FAIR原则的基础上, 针对本土数据提出了CARE原则^[39]; 针对科研软件共享, 提出了FAIR4RS原则^[40]; 针对提升科学数据可用性, 沈志宏等人^[41]提出PARIS原则(Processable, Askable, Reliable, Incorporable and Suppliable), 可实现科学数据从“可用”到“可协作”, 丰富了数据可用性的内涵.

科学数据的管理贯穿科学数据的全生命周期, 中国科学院经过30多年的实践积累, 构建了科学数据标准体系^[42], 涵盖数据采集与整理、元数据与数据模型、系统与接口、数据管理、数据质量以及数据服务

等方面. 此外, 中国国家科技基础条件平台在2018年提出了一套科学数据标准规范体系^[43], 全国科技平台标准化技术委员从科技资源类型和全生命周期管理的维度, 提出了科技资源管理体系框架^[44].

科学数据的元数据标准规范研究是数据可理解性、人工和机器友好性及具备可互操作能力的重要保障, 无论是国际标准化组织(International Organization for Standardization, ISO)、中国国家标准还是行业标准都制定出台了科学数据相关系列标准, 内容涉及描述性元数据、元数据管理、元数据注册与索引等^[45], 覆盖通用型的标准和具体专业领域的标准. 运用较为广泛的通用型的元数据标准包括都柏林核心(Dublin Core, DC)、DataCite Metadata Schema(下文简称DataCite)、Schema.org(<https://schema.org/>)等. 比较而言, DC元数据(2012年版)包含15个元素^[46], DataCite 4.4^[47]含20个元素, 两者有12个重合元素(见表2); Schema.org则由Google、Microsoft、Yahoo和Yandex公司联合创建, 在搜索引擎可发现性上有更广泛的应用支持.

聚焦人口神经科学, 研究人员对神经影像大数据具有学科内在需求, 因此在神经影像数据的共享上, OHBM倡导神经影像数据及代码共享的最佳实践以提

表 2 Dublin Core与DataCite Metadata Schema 4.4元数据元素^{a)}

Table 2 Dublin Core and DataCite Metadata Schema 4.4 metadata elements

元数据标准	元数据元素集
Dublin Core	<u>title</u> 、 <u>creator</u> 、 <u>contributor</u> 、 <u>identifier</u> 、 <u>language</u> 、 <u>format</u> 、 <u>subject</u> 、 <u>type</u> 、 <u>description</u> 、 <u>date</u> 、 <u>publisher</u> 、 <u>coverage</u> 、 <u>rights</u> 、 <u>relation</u> 、 <u>source</u>
DataCite Metadata Schema 4.4	<u>Title</u> *、 <u>Creator</u> *、 <u>Contributor</u> 、 <u>Identifier</u> *、 <u>Language</u> 、 <u>Format</u> 、 <u>Subject</u> 、 <u>ResourceType</u> *、 <u>Description</u> 、 <u>Date</u> 、 <u>PublicationYear</u> 、 <u>Publisher</u> 、 <u>Rights</u> 、 <u>RelatedIdentifier</u> 、 <u>AlternateIdentifier</u> 、 <u>RelatedItem</u> 、 <u>GeoLocation</u> 、 <u>FundingReference</u> 、 <u>Size</u> 、 <u>Version</u>

a) *项为DataCite Metadata Schema 4.4标准中的必备元数据元素项; 带下划线的元素为DC和DataCite的重合元素项

升研究可复现性和可复制性, 并积极开展探索工作. 2016年, OHBM推出COBIDAS MRI报告^[48], 2018年推出COBIDAS MEEG报告, 且随着MEEG硬件、软件和方法的发展及迅速变化, 该报告持续动态更新^[49]. 此外, 国际神经信息学协调机构(The International Neuroinformatics Coordinating Facility, INCF)致力于协调基础设施和标准以促进数据共享, 为支持FAIR原则, 启动了对社区标准及最佳实践工作的评估和认定^[50], 在领域标准框架设计和扩展现有社区标准的框架设计上开展工作. 在数据格式上也形成了社区标准, 如脑成像数据结构(Brain Imaging Data Structure, BIDS)统一了领域内常见的数据格式, 覆盖了常见数据处理操作所需的元数据, 使得数据共享和重用成为可能^[51]; 为便利BIDS标准的推广实践, 社群提供了BIDS验证器(BIDS validator)^[52]供研究人员检查数据是否符合标准要求.

3.2 建设数据共享支撑平台的原则与指南

从审计数据仓储库的可信服务能力出发, 世界数据系统(World Data System, WDS)提供CoreTrustSeal^[53](CTS)认证服务, 在组织建设能力、数字资源管理、信息技术与安全管理三个维度的16项细则作出评估认证. 此外, 还包括有DIN31466/NESTOR^[54], ISO16363^[55]等标准. 从资金资助方和数据用户的需求角度出发, 国际组织研究数据联盟(Research Data Alliance, RDA)提出数据仓储库的TRUST原则(Transparency、Responsibility、User focus、Sustainability、Technology)^[56].

在政府层面, 美国国家科学技术委员会(National Science and Technology Council)提出“联邦资助研究数据存储库应具备的特性列表”(https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf), NIH(https://sharing.nih.gov/data-management-and-sharing-policy/

sharing-scientific-data/selecting-a-data-repository#desirable-characteristics-for-all-data-repositories)和NSF在该框架下对其资助数据的共享存储制定政策要求, 除个别表述及排列顺序上存在差异, 基本沿用CTS的三个评价维度; 但针对涉及人类受试者的数据存储, 提出额外7项要求, 即是数据的访问和使用完全遵循参与者的同意范围(Fidelity to Consent), 实施隐私保护并提供措施文档(Security), 可遵守实施数据使用限制(Limited Use Compliant), 利用已有透明的流程审查数据访问请求(Request Review), 安全措施检测数据泄漏及响应(Plan for Break), 可控制和审核数据的访问和下载(Download Control), 制定了违规行为条款和处理程序(Accountability).

出版商则更需要数据存储库提供开放可无障碍访问的数据获取服务, 且满足论文出版的相关需求. 在RDA工作组提出的期刊和出版商数据政策框架中^[57], 涵盖了对数据开展同行评议的需求; 在STM提出的“数据共享的12个最佳实践”(https://www.stm-assoc.org/research-data-program/bestpractices/)中, 数据存储库要确保实践数据的FAIR原则, 确保建立论文与数据之间的连接关系, 通过系统间协作确保数据能够数据轻松、清晰、有效共享. 出版商结合自身数据政策, 也提出了不同要求, 如需存储库提供开放许可协议(如CC0、CC-BY等), 数据访问模式可支持同行评议需要等.

3.3 神经科学领域数据存储服务平台的建设实践

据统计, 截至2024年1月, re3data(https://www.re3-data.org/)上注册的数据存储库数量为3182个, 其中神经科学(Neuroscience)领域的注册数据存储库数量为113个, 主要集中在北美及欧洲国家(见图2). 这一方面说明神经科学领域在科学数据共享和相关信息化服务平台建设上的前瞻性, 体现了其在数据共享方面的实

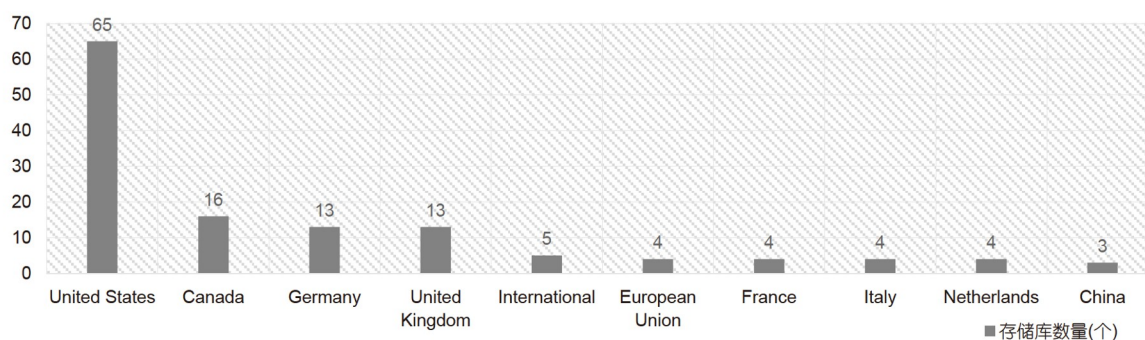


图 2 在re3data注册的神经科学领域数据存储库的国家分布情况统计(数量排名前十)

Figure 2 Statistics on the national distribution of neuroscience data repositories registered at re3data (top 10 by number)

践已具良好基础，另一方面说明经济体的发展水平在科学数据共享建设实践上呈现出不均衡性和多样性。

聚焦神经科学领域的数据共享平台建设，除前文已提到的HCP^[16]、FCP1000^[15]、UK Biobank^[18]、ABCD^[17]和PPMI^[36]外，国际上还有OpenNeuro（前称OpenfMRI）^[58]、CoRR(Consortium for Reliability and Reproducibility)^[59]、PNC (Philadelphia Neurodevelopmental Cohort)^[60]、NDAR (National Database for Autism Research)^[61]、ABIDE(Autism Brain Imaging Data Exchange)^[62]、Healthy Brain Network(HBN)^[63]、Cambridge Centre for Ageing and Neuroscience(Cam-CAN)^[64]等神经影像数据库(表3)。这些数据库的存在和发展不仅促进了脑影像学和相关领域的研究，还加强了全球科学界在数据共享和合作方面的努力。

在我国科学数据开放共享政策的大背景下，认知神经科学家积极响应号召，通过个体层面积极推动数据的开放共享，让科学数据公开成为了可能。左西年等人基于国家基础科学数据中心“活体人口影像交叉学科脑库”，发布了“中国人彩巢计划(Chinese Color Nest Project, CCNP)”大脑影像数据，推动发展认知神经科学向群体迈进。该项目依托脑科学与大数据的科学研究成果，推进服务基础教育阶段学龄儿童青少年的人才培养，响应国家儿童青少年脑智开发和心理健康重大需求，有效连接心理学、认知神经科学和教育学三大领域，助力国家基础教育高质量发展。项目的数据资源向全球研究者开放。通过提供便捷的数据访问接口和丰富的数据集下载服务，为中国及全球脑科学研究提供了重要的基础设施支持。

2023年，“中国人彩巢计划”项目委员会和科学数据银行(Science Data Bank, ScienceDB)结合国际社区

与国内社区的开放科学文化，合作共建交叉学科数据共享合作的公共平台“彩巢计划-中国人脑智毕生发展数据社区”(https://ccnp.scidb.cn, 下简称ScienceDB·CCNP)，平台专注于脑影像数据共享，打造高质量中国人脑智毕生发展数据共享环境，促进科研合作、知识传播和数据FAIR共享，打破地域和机构的界限，提升研究的透明度和可信度及其成果价值和影响力。减少重复研究和数据采集的成本，提高科研资金的利用效率。依托科学数据银行的基础设施服务能力和配套管理体系，平台具备安全可靠的服务能力，在数据共享方面形成了独特的服务特色(表4)。

ScienceDB·CCNP不仅收集了不同背景个体的脑成像数据，更为跨学科间合作交叉创新提供了平台，让计算机科学家、心理学家、神经科学家和医学研究人员能够共同研究大脑的奥秘。ScienceDB·CCNP的建设依托平台ScienceDB作为通用型科学数据存储库，与各学科领域合作建设形成多学科的领域数据社区，如与国家空间科学数据中心共建的ScienceDB·Space社区，与国家生态科学数据中心共建的ScienceDB·EcoDB社区，与国家农业科学数据中心共建的ScienceDB·Agriculture&Rural Area社区，与中华医学会杂志社共建的ScienceDB·CMA等。ScienceDB·CCNP与ScienceDB上的多领域科学数据遵循统一的核心元数据标准，具有友好的机器可读性，为跨学科融合研究提供基础服务保障。“中国人彩巢计划”的价值，不仅在于数据的丰富度和开放性，更在于它所倡导的数据共享和科学交流文化。这种文化促使数据、方法、工具和见解的分享，加强了科研社区内的合作和创新能力。

总体上，开放式神经影像的发展历程反映了科学研究从孤立的个体研究向协作共享的转变，同时，现有

表 3 神经科学领域部分国际数据库情况简介

Table 3 Overview of some international databases in neuroscience

数据库/数据集名称	资助/建设方	开放数据情况简介
OpenNeuro (前称OpenfMRI) ^[58]	美国国家健康研究院(National Institutes of Health, NIH)资助, 斯坦福大学Poldrack实验室运营	该平台旨在通过提供高质量的神经影像数据, 促进神经科学研究的透明性和可重复性. OpenNeuro汇集了来自全球多个研究项目的功能磁共振成像(functional magnetic resonance imaging, fMRI)、结构磁共振成像(structural magnetic resonance imaging, sMRI)、扩散加权成像(diffusion weighted imaging, DWI)等多模态数据, 涵盖认知、情感、行为等多个研究领域. 研究者可以自由访问和上传数据, 进行各类神经科学研究. 其数据管理机制确保了数据的准确性和可靠性, 其开放共享的理念极大地推动了神经科学研究的发展, 促进了国际科研合作和学术交流
国际信度与可重复性联盟 (Consortium for Reliability and Reproducibility, CoRR) ^[59]	美国国家药物滥用研究所(National Institute on Drug Abuse, NIDA)和国家自然科学基金会(National Natural Science Foundation of China, NSFC)在CoRR合作中发挥了重要作用, 儿童心理研究所、中国科学院心理研究所和内森克莱恩研究所一起提供了必要的资金和人力来构建该项目的基 础	收集和共享多中心的高质量神经影像数据, 致力于建立标准化的数据收集和分析方法. 该联盟汇集了来自多个国家和地区的研究机构, 涵盖了大量sMRI、fMRI和DWI数据. 其数据资源为全球研究者提供了用于验证和重复已有的研究结果的平台, 提高了神经科学研究的可信度和科学性. 在神经科学研究 的透明性和可重复性方面发挥了重要作用
费城神经发育队列 (Philadelphia Neurodevelopmental Cohort, PNC) ^[60]	由宾夕法尼亚大学佩雷尔曼医学院与 儿童医院合作开展的大型研究项目	该项目旨在通过多模态神经影像学技术研究青少年时期的大脑发育和心理健康. PNC收集了约9000名8~21岁参与者的广泛数据, 包括sMRI、fMRI、DWI、认知测试和精神健康评估数据. PNC的数据资源向全球研究者开放, 支持各类脑发育和心理健康研究. 通过PNC的数据共享机制, 研究者能够深入探讨青少年大脑发育的规律及其与心理健康问题的关系, 从而为相关的诊断和治疗提供科学依据
国家自闭症研究数据库 (National Database for Autism Research, NDAR) ^[61]	NIH资助的一个专注于自闭症谱系 障碍研究的大型数据平台	NDAR汇集了来自多个研究项目和机构的大量自闭症相关数据, 包括基因组数据、神经影像数据、行为和临床评估数据等. 该数据库旨在通过标准化和共享自闭症研究数据, 促进自闭症的诊断、治疗和机制研究. 研究者可以自由访问和上传数据, 进行各类自闭症研究. NDAR的数据管理和共享机制确保了数据的高质量和可重复性, 极大地推动了自闭症研究的进展, 为了解和应对自闭症提供了宝贵的资源
自闭症脑影像数据集(Autism Brain Imaging Data Exchange, ABIDE) ^[62]	主要由美国国立精神健康研究所 (National Institute of Mental Health, NIMH)、Leon Levy基金会等资助	收集和共享来自全球多中心的神经影像数据, 致力于提高自闭症研究的数据一致性和可重复性. 其开放共享机制极大地推动了自闭症研究的进展, 为自闭症的诊断、治疗和机制神经生物学机制及其与行为关系研究提供了重要的科学依据
脑健康数据集(Healthy Brain Network, HBN) ^[63]	纽约儿童心理健康研究基金会(Child Mind Institute, CMI)发起的开创性 研究项目	旨在通过多模态数据的收集和共享, 深入理解儿童和青少年的脑健康状况. HBN项目收集了广泛的神经影像数据、遗传学数据、心理和行为评估数据, 以及认知功能测试数据, 涵盖了5~21岁年龄段的多样化人群. 促进儿童和青少年脑健康研究的透明性和可重复性. 研究者可以自由访问和使用HBN的数据, 以探讨大脑结构和功能的发育规律及其与心理健康问题的关系. HBN为研究儿童和青少年心理健康问题提供了宝贵的科学依据, 在全球范围内促进了科学合作和知识共享, 为脑健康研究提供了重要支持
剑桥老龄化和神经科学数据集(Cambridge Centre for Ageing and Neuroscience, Cam-CAN) 项目 ^[64]	由英国由生物技术和生物科学研究委员会 (Biotechnology and Biological Sciences Research Council, BBSRC)提供大量资金 建设支持	目涉及3000多名18~88岁参与者, 涵盖了不同年龄段和健康状况的人群, 为研究大脑老化的生物学机制提供了丰富的数据资源. Cam-CAN的数据资源对全球研究者开放, 支持各类与老龄化相关的神经科学研究. 通过促进数据共享和跨学科合作, Cam-CAN在提高老龄化研究的透明性和可重复性方面发挥了重要作用, 为延缓老龄化相关的认知衰退和疾病提供了科学依据和潜在的干预策略

表 4 ScienceDB-CCNP特色服务情况一览介绍
Table 4 Overview of ScienceDB-CCNP featured services

特色服务	服务简介
不同粒度的数据共享支持	除了以数据集(dataset)为单位进行数据共享, 还能够提供以数据包(data package)和数据文件(data file)粒度的数据共享. 各粒度的数据均分配有唯一标识符(CSTR或DOI), 可被独立传播、独立引用和独立计量
支持数据文件目录及数据文件的可视化预览	平台支持大体量(TB级)的数据共享, 支持以文件和文件夹目录的形式共享数据, 且提供文件夹的在线浏览、文件的在线检索、2000多种常用数据文件格式的在线预览服务等, 可确保数据的便捷获取与访问
提供数据使用情况追踪统计	通过永久标识符(CSTR 和 DOI)系统追踪每个数据包、数据集和数据文件的引用情况, 使用情况(访问、下载及其地理分布、社交关注度等情况), 收集用户点赞、收藏等行为数据, 全方面反映数据使用情况
联通数据知识产权登记	在激励数据共享行为上, ScienceDB-CCNP可链接数据知识产权登记平台. 数据作者完成数据提交且通过审核后, 可申请进行数据知识产权登记工作. 登记审理工作由提交至的知识产权登记机构进行审核与发证
遵循国际标准 联通国际资源体系	平台数据均配有明细的数据使用许可协议(如CC-BY等), 确保数据可重用. 此外, 支撑平台ScienceDB满足支持多项国际标准, 被写入多家国际出版商(如Springer Nature、Cell Press、Elsevier、Taylor & Francis等)的推荐存储库列表, 并被多家国际科学数据索引平台(如Data Citation Index、Scholix、Google Dataset Search、Dimensions等)收录, 进行全球传播

表 5 现有脑影像数据库特点归纳
Table 5 Summary of characteristics of existing brain imaging databases

特点	特点描述
大数据集成	大规模的数据集是这些数据库的一个显著特点. 这些集成的大数据集使得研究者可以进行全面、深入的分析, 尤其是在探索罕见现象或微妙的神经影像差异方面
标准化的数据处理	为了确保数据质量和分析的一致性, 许多数据库采用标准化的数据处理和分析指南
开放存取和透明度	许多数据库遵循开放存取原则, 允许研究者共享他们的发现, 促进了知识的广泛传播和后续研究
多样性与专业化	不同的数据库可能关注于不同的研究主题, 如特定类型的神经疾病、特定人群的大脑发育, 或特定的神经成像技术

脑影像数据库呈现出大数据集成、标准化数据处理、具备开放存取与透明度、多样性与专业化的特征(见表5). 通过构建系统化、标准化的开放数据库, 全球研究者能够更加便捷地获取和利用丰富的神经影像数据, 从而推动了神经科学大数据及可重复研究的进步. 这些开放数据库不仅在学术研究上具有重要意义, 还为临床诊断和治疗提供了宝贵的参考, 为推动脑科学的全面发展作出了巨大贡献.

4 数据开放共享可持续发展的问题与挑战

4.1 人类相关科学数据的伦理与安全问题挑战

数据伦理与安全是项重要的议题, 在数据共享的操作过程中也一直面临着巨大挑战. 在神经科学领域, 尤其是来自人类受试者的数据, 对伦理审核机制提出

了更高要求. 伦理审核的目的在于保证研究符合道德标准, 尊重并保护参与者的权利和隐私. 这需要配套透明、严格、高效的审核流程. 但在处理跨国界、多机构的大规模数据共享场景时, 伦理标准差异, 往往也伴随着隐私安全法规上的差异, 时常成为阻碍. 因此, 建立广泛接受的国际伦理准则, 协调地区政治文化差异成为挑战. 同时, 新技术的发展引入了新的伦理问题, 伦理审核标准也需不断更新以应对变化. 例如, 脑机接口和人工智能在认知神经科学中的应用就引发了关于人类自主权、数据安全和算法透明度的伦理问题. 因此, 建立动态、灵活的伦理审核机制以保障科研的道德性和合法性是另一重大挑战.

认知神经科学收集的数据通常包括个体的脑活动模式、心理测试结果和生物标记物等敏感信息. 这些数据对于科学发现和医学进步至关重要, 却存在着隐

私泄露和不当使用的风险。从个体隐私安全的层面,欧盟、美国及中国已开展了诸多立法工作,如欧盟于2018年生效的“世上最严格”数据保护法案《一般数据保护条例》,美国1996年颁布的《健康保险流通与责任法案》,我国先后颁布的《网络安全法》、《数据安全法》、《个人信息保护法》等。这些法规都保护隐私信息和敏感数据,但实施方式和保护范围存在巨大差异^[65]。此外,在实践层面,多数神经科学领域科研人员(45.2%)表示如果进行数据共享,将无法确保数据匿名性^[66]。随着AI技术发展,面部重建技术的去识别方法失效^[67]等新问题不断出现^[68]。我国虽先后出台了《新一代人工智能伦理规范》,《生成式人工智能服务管理暂行办法》以应对新的伦理与安全挑战,但平衡数据共享与隐私安全仍将会是个长期而复杂的问题。

在保护国家生物信息安全的层面,2024年3月国家网络安全和信息化委员会办公室发布《促进和规范数据跨境流动规定》(https://www.cac.gov.cn/2024-03/22/c_1712776611775634.htm)提出学术合作等活动中收集和产生的数据,在不包含个人信息或重要数据的情况下,可免于申报出境。但对于敏感数据,2024年2月28日,拜登政府颁布《阻止有关国家获取美国公民大规模敏感个人数据和美国政府相关数据的行政命令》(<https://www.whitehouse.gov/briefing-room/presidential-actions/2024/02/28/executive-order-on-preventing-access-to-americans-bulk-sensitive-personal-data-and-united-states-government-related-data-by-countries-of-concern/>),重点关注包括基因组数据、生物识别数据、个人健康数据、地理位置数据等敏感数据,首次提出限制向包括中国在内的出售禁令^[69]。《促进和规范数据跨境流动规定》虽对学术合作中的非敏数据跨境流动开放了通路,但对我国“重要数据”分级的可实践路径及责任落实机制目前尚缺顶层设计和推进手段,数据跨境流通的安全性仍存在诸多不确定因素,极不利于稳健的数据共享环境建设和长期的国际合作开展。

4.2 数据共享激励机制与学术认可不足

“获得的认可不足”始终都是科研人员在数据共享上面临的突出问题^[70]。一项2022年的调查显示^[66],在神经科学领域有11%的受访者从未考虑过共享数据。具体到数据管理与数据共享的阻碍因素,各种调查结果^[66,70,71]都高度一致地集中在“数据整理工作过于耗

时”“耗资”“害怕成果被抢发”等原因,其根源还是来自数据共享行为无法获得足够的认可。

发表数据论文是现阶段被公认为最有效的实践手段。但在全球数据期刊的数量(表1)和数据论文的发表量上,总体占比仍非常小。据Web of Science收录数据显示,2023年出版论文量约为304.6万篇,数据论文出版量为2269篇,占比仅为0.74%(图3统计了2023年Web of Science上收录数据论文的学科分布情况)。在鼓励期刊开展数据论文出版实践,引导学者发表数据论文上,仍需更多政策引导和配套标准的推出实施。同时,在科研群体内普及和鼓励规范的数据引用十分重要^[72]。Force11工作组曾于2014年发表“数据引用原则联合声明”(Joint Declaration of Data Citation Principles, JDDCCP)^[73],并相继提出出版商^[74]、数据存储库^[75]的实施路径图,最佳实践的广泛普及存在较高提升空间。

在其他正向激励机制上,如获得资金资助、晋升机会等,已有研究探索和路线图(如欧盟委员会资助开展了FAIRReR评估研究,期将FAIR数据纳入欧洲研究评估的研究, <https://avointiede.fi/en/networks/eosc-co-creation>),但文化与社会认同上的挑战以及配套基础设施能力的建设仍阻碍重重。

4.3 系统化的科学数据管理能力培育缺失

随着更多政策及法规出台,面向科研社群的相关培训和支撑缺失问题愈发凸显。一项来自2023年全球问卷统计显示,有近四分之三的科研工作者从未接受过数据计划、管理及共享上的支持;有45%的生物科学领域人员表示在“寻找合适的数据仓储库”上存在困难^[70];类似地,在一项2022年开展的面向神经影像学数据共享的调查研究显示^[66],41%的受访者对合适的存储库知之甚少,40%的受访者从来没有接受过关于如何在线共享数据的培训。事实上,在政府的、基金资助方的还是出版商的数据政策中,绝大部分都专设了关于数据存储库遴选推荐的内容,相关政策支撑和能力培育能力的可提升空间可见一斑。

日趋严格和健全的数据共享政策对于提升科研共同体对数据共享的认知与实践参与无疑是种有效手段,并在推动科学研究的透明度和协作方面发挥着重要作用。但对研究人员而言,数据共享需要满足国家法律法规、项目资助方、高校或研究所、期刊出版商、科研共同体等的要求,必要及完备的指导及标准存在缺失,这会使得管理负担日益加重、政策认同感降低并难以

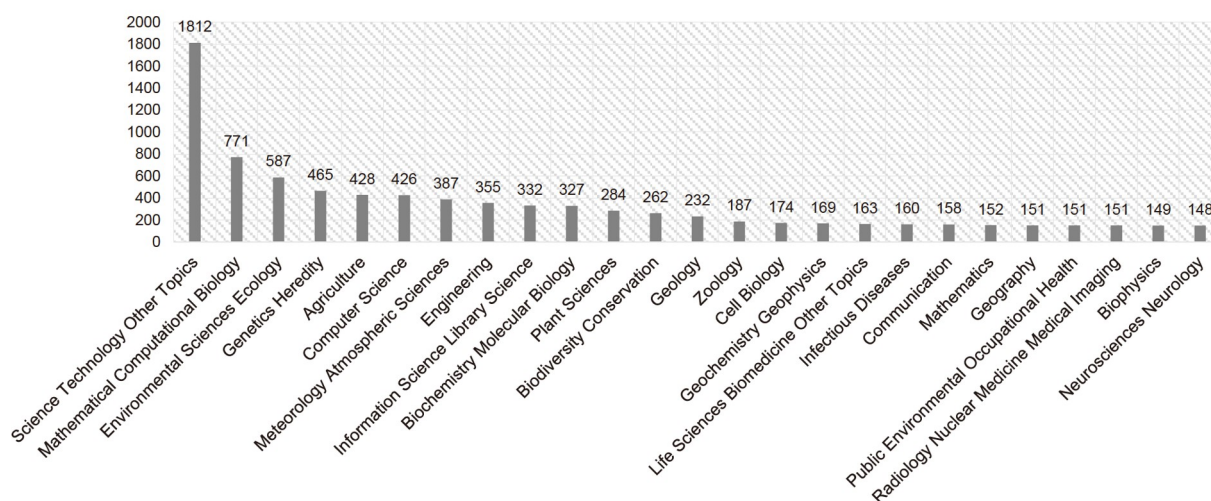


图3 Web of Science 2023年出版数据论文的学科分布情况统计(数量排名前25的学科)

Figure 3 Distribution of disciplines for data papers published in 2023 in the Web of Science (Top 25 disciplines by number)

付诸实施。

5 思考与建议

5.1 加强顶层布局, 集聚科研共同体的力量创新数据共享形式

国际范围内, 科学数据的开放共享已成为重要共识。开放共享与安全的平衡是一个随着科技发展、政治环境而不断变化的问题, 如何最大化地实现发展, 并确保个人和国家安全, 需要政府层面整体谋划, 通过顶层设计引导共享环境良性健康发展, 并根据发展阶段和技术发展水平变化, 及时做出更新调整。神经影像学是一个从数据共享中获益匪浅的科学领域^[76], 大量神经科学数据正随着世界各地发起的研究计划不断产生, 如何规避安全风险开展国际合作, 如何利用新技术创新数据共享机制, 促进数据效能释放, 推进学科发展与跨领域创新发现, 仍需探索实践。在自下而上的自驱实践中, 须集聚跨学科领域的专家共同参与, 在伦理与隐私安全等关键复杂问题上寻找有效实践路径。

在数据共享的标准规范建设上, FAIR原则虽提出了数据共享的最佳实践原则, 但在具体实现数据共享的可互操作性(Interoperability)和可重用性(Reusability)原则仍要科研社区制定细则, 这需要科研共同体的共同参与。聚焦神经科学, 高昂的数据获取成本需要数据共享来实现可重用的治理目标, 领域元数据内容除了FAIR原则的建议外, 还包括有人类受试者的描述、实

验、实验数据的性质、附加测试和调查, 以及知情同意及使用限制等内容^[67], 通过标准化的事件注释, 神经影像数据集可用于项目设计外的研究用途^[77]。相关国际组织需在制定行业数据共享规范上提出更多具有普遍性的实践建议, 如以提升数据可重用性为目标(包括对人和对机器的可读性)的元数据标准等标准体系并配套推广应用。此外, 也呼吁我国科学家更多地参与到领域数据治理中, 在引领国际标准制定、国际组织数据政策等方面发挥更多主导作用。

5.2 推进数据共享激励机制研究与学术贡献认同, 激发科研人员数据共享自驱力

良性的数据共享生态需要所有贡献者都能获得与投入相匹配的收益, 需要配套的数据价值评估机制、激励机制、学术认同环境以及创新的技术保障, 是确保政策得以推进实施, 并朝着数据共享的初衷而非流于形式推进的重要保证。

一是建立数据价值评估机制。实践表明, 并非所有的数据共享都能带来研究的可复现性^[78]。这需要健全数据标准体系建设, 形成科学、客观的数据价值评估机制并配套政策引导, 从而实现提高数据可用性和重用的目标。二是不断完善激励机制。建立多元化的激励机制措施, 如资金奖励、项目资助优先权、学术评价等; 将数据共享纳入绩效评估, 确保数据共享可得到相应的认可、奖励或晋升机会; 完善政策法规与知识产权保护, 使得数据共享获法律上的保护。三是强化对数

据共享者学术贡献的认同. 推广数据引用和署名, 确保数据贡献者得到充分认可和标引, 基础设施服务配套做好成果标识和贡献者标识服务, 确保贡献信息的可关联追溯能力^[79]; 健全学术评价指标, 更新现有以论文为中心的评价模式, 引入数据共享贡献指标, 补充传统的h指数等学术评价体系; 设立相关奖项, 提升贡献者学术声誉. 四是开展创新模式设计与平台技术保障建设, 实现投入与回报相匹配的激励机制模式, 如利用区块链等技术手段, 让数据共享获取积分换来更多高价值或稀缺数据、经济回报等机制.

5.3 高质量的数据共享服务建设促进数据共享实践

在各种数据政策中, 数据存储政策都是重要的组成部分, 承载数据存储与获取工作的基础设施及工具的服务能力, 对确保科研人员实施政策要求, 低成本实践数据共享都有重要影响. 据图2所示数据仓储服务平台的国家分布情况来看, 此外, 我国在数据管理、仓储和服务平台上的建设投入与欧美西方国家相比差距十分明显, 这与我国是数据生产大国, 同时也是数据贡献大国的发展现状严重不匹配. 我国政府与资金资助方须加大数据共享服务平台的投入, 这即是有效管理和长效积累公共财政资助产出数据的需要, 也是服务我国科学家提升科研效率与质量的需要, 更是作为一个科研产出大国主导国际合作, 获得学术话语权的需要.

高质量的数据共享服务可提供高效的海量数据处理能力. 在认知神经科学领域, 研究人员需要处理大量的脑成像数据, 强大的基础设施可解决数据存储、管理及索引上的刚需. 其次, 高质量的数据共享基础设施应提供数据标准化工具. 不同的测量和分析技术会产生不同的数据格式, 标准化工具可协助数据格式转换以便跨实验室、跨学科的数据共享和比较. 此外, 高质量的数据共享基础设施应包含数据分析和可视化工具, 辅助数据分析, 促进新模式和新联系的发现. 如, 脑网络分析工具可以帮助探索大脑不同区域间的连接和交互, 从而对大脑功能有更深入的理解. 最后, 高质量的数据共享基础设施须提供培训和支持服务. 认知神经科学领域的研究技术和分析方法日新月异, 关于数据共享和分析工具在线教程、研讨会和技术支持十分必要的.

5.4 创造跨学科融合与交叉创新机会

科学研究开启“第五范式”——智能化科研(AI for Science, AI4S)时期, 并将逐步成为科学研究的主流范式之一^[80]. AI4S正在为认知神经科学领域研究带来变革. 机器学习和深度学习等技术已被广泛应用于处理和分析庞大的脑成像数据集, 在识别大脑活动中的复杂模式, 在疾病的早期阶段发现, 提高诊断准确性和效率等方面提供有效帮助. 在模拟大脑功能, 揭示神经活动基本机制, 研究大脑处理信息、控制行为等机制问题上AI也发挥着重要作用. 此外, 根据海量文献和数据, AI能够提出新的研究假设和方法论建议, 大大提高研究效率和质量.

此外, AI4S的重要特征是“平台型科研”, 数据、算法模型和算力缺一不可^[80]. 也有学者提出顶层设计要将基础设施建设放在第1位^[81]. 以“中国人彩巢计划”脑影像数据库为例的平台展示了跨学科合作的巨大潜力和AI在促进科学研究中的关键作用. 支持和维护这样的科研基础设施, 对于推进科学整体进步至关重要. 通过加强不同学科之间的合作, 促进知识和技术的交流, 我们不仅能深入理解大脑等复杂系统, 还能借助AI等前沿技术, 创造更多跨学科融合与交叉创新的机会.

5.5 不同文化差异下的开放数据实践

在不同的文化背景下, 开放数据的实施路径呈现出不同特征. 在神经成像领域与心理学领域, 西方背景下的更多实践是自下而上的发展模式, 很多情况下科研工作者仅靠个人或团队无法获得所有数据, 数据共享使得有限的资金条件可获得更大规模的研究样本^[82], 实现突破性进展从而获得更多的来自社会公众与基金会的资金支持, 推进了来自政府、资金资助方的政策支持和要求, 其中, BIDS标准就是典型的科研社群自发驱动实践的产物. 在国内, 许多基础设施的建设则具有自上而下的特点, 而科研群体中的共享文化氛围存在欠缺. 不同的开放数据路径具有各自的优势和现实挑战, 存在彼此借鉴的参考意义, 也须发挥建制化优势, 激发我国科研群体自下而上地投入开放数据实践与治理, 如何探索出适合国内文化和社会特点的开放数据路径值得思考.

参考文献

- 1 Zhang Q, Wang Y S, Chen L Z, et al. Normative modeling for developmental population neuroscience: A “microscope” through which the laws and

- characteristics of individual differentiation can be quantified in human brain-mind development (in Chinese). *Chin Sci Bull*, 2023, 68: 2086–2100 [张青, 王银山, 陈丽珍, 等. 发展人口神经科学中的规范化建模: 量化脑智发展规律与个体分化特征的“显微镜”. *科学通报*, 2023, 68: 2086–2100]
- 2 Gong Z Q, Biswal B B, Zuo X N. Paradigm shift in psychiatric neuroscience: Multidimensional integrative theory. *Med Plus*, 2024, 1: 100024
 - 3 Zhou Z X, Chen L Z, Milham M P, et al. Six cornerstones for translational brain charts. *Sci Bull*, 2023, 68: 795–799
 - 4 Lu B, Chen X, Xavier Castellanos F, et al. The power of many brains: Catalyzing neuropsychiatric discovery through open neuroimaging data and large-scale collaboration. *Sci Bull*, 2024, 69: 1536–1555
 - 5 Chambers C. *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton: Princeton University Press, 2017
 - 6 Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 2015, 349: aac4716
 - 7 Marek S, Tervo-Clemmens B, Calabro F J, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature*, 2022, 603: 654–660
 - 8 Anvari F, Lakens D. The replicability crisis and public trust in psychological science. *Compr Results Soc Psychol*, 2018, 3: 266–286
 - 9 Botvinik-Nezer R, Wager T D. Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biol Psychiatry-Cogn Neurosci NeuroImag*, 2023, 8: 780–788
 - 10 Hu C, Wang F, Guo J, et al. The replication crisis in psychological research (in Chinese). *Adv Psychol Sci*, 2016, 24: 1504–1518 [胡传鹏, 王非, 宋梦迪, 等. 心理学研究中的可重复性问题: 从危机到契机. *心理科学进展*, 2016, 24: 1504–1518]
 - 11 Goodman S N, Fanelli D, Ioannidis J P A. What does research reproducibility mean? *Sci Transl Med*, 2016, 8: 341ps12
 - 12 Milham M P, Craddock R C, Son J J, et al. Assessment of the impact of shared brain imaging data on the scientific literature. *Nat Commun*, 2018, 9: 2818
 - 13 Poldrack R A, Gorgolewski K J. Making big data open: Data sharing in neuroimaging. *Nat Neurosci*, 2014, 17: 1510–1517
 - 14 Emanuele E, Minoretti P. Measuring the Impact of Data Sharing: From Author-Level Metrics to Quantification of Economic and Non-tangible Benefits. *Cureus*, 2023, doi: 10.7759/cureus.50308
 - 15 Biswal B B, Mennes M, Zuo X N, et al. Toward discovery science of human brain function. *Proc Natl Acad Sci USA*, 2010, 107: 4734–4739
 - 16 Marcus D S, Harms M P, Snyder A Z, et al. Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage*, 2013, 80: 202–219
 - 17 Jernigan T L, Brown S A, Dowling G J. The adolescent brain cognitive development study. *J Res Adolesc*, 2018, 28: 154–156
 - 18 Miller K L, Alfaro-Almagro F, Bangerter N K, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*, 2016, 19: 1523–1536
 - 19 Brakewood B, Poldrack R A. The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage*, 2013, 82: 671–676
 - 20 Kozlov M. NIH issues a seismic mandate: Share data publicly. *Nature*, 2022, 602: 558–559
 - 21 Zhang Y N, Zhang C H, Kang J, et al. Policies and planning for advancing scientific data development in Australia (in Chinese). *China Sci Data*, 2024, 9: 36–51 [张耀南, 张彩荷, 康建芳, 等. 澳大利亚科学数据发展政策与规划梳理. *中国科学数据*, 2024, 9: 36–51]
 - 22 Zhang Y N, Zhang J, Ren Z Y, et al. Policies and planning for advancing scientific data development in BRICS (in Chinese). *China Sci Data*, 2024, 9: 52–73 [张耀南, 张杰, 任泽瑶, 等. 金砖国家科学数据发展政策与规划. *中国科学数据*, 2024, 9: 52–73]
 - 23 Thorp H H, Vinson V, Yeston J. Strengthening the scientific record. *Science*, 2023, 380: 13
 - 24 Zhao R X, Zhao H, Zhu L. Progress in the development and sharing of big data in agricultural science between china and foreign countries (in Chinese). *J Agric Big Data*, 2019, 1: 24–37 [赵瑞雪, 赵华, 朱亮. 国内外农业科学大数据建设与共享进展. *农业大数据学报*, 2019, 1: 24–37]
 - 25 Molodtsov F, Nikiforova A. From an integrated usability framework to lessons on usability and performance of open government data portals: A comparative study of European Union and Gulf Cooperation Council Countries. arXiv preprint, arXiv: 2406.08774, 2024
 - 26 Boulton G, Campbell P, FREng B C, et al. *Science as An Open Enterprise*. London: The Royal Society Science Policy Centre, 2012
 - 27 Zhang Y N, Ren Z Y, Kang J F, et al. Policies and planning for advancing scientific data development in the UK (in Chinese). *China Sci Data*, 2024, 9: 21–35 [张耀南, 任泽瑶, 康建芳, 等. 英国科学数据发展政策与规划梳理. *中国科学数据*, 2024, 9: 21–35]
 - 28 Vrouwenvelder K, Stall S, Giampaola M. Request for Information (RFI) on NSF Public Access Plan 2.0: Ensuring Open, Immediate, and Equitable

- Access to National Science Foundation Funded Research. Authorea Preprints, 2024
- 29 Wang R D, Gao M X, Shi L, et al. Research and thoughts on the opening and sharing of scientific data under background of Big Data (in Chinese). *China Sci Technol Res Rev*, 2020, 52: 1–5, 26 [王瑞丹, 高孟绪, 石蕾, 等. 对大数据背景下科学数据开放共享的研究与思考. *中国科技资源导刊*, 2020, 52: 1–5, 26]
- 30 Smit E, van Rossum J. STM's Research Data Year: Results, lessons learnt, and next steps. *Learned Publish*, 2022, 35: 67–70
- 31 Peng L, Han Y L. Analysis and enlightenment of the data policy of Chinese scientific journals: Taking the English journals sponsored by Chinese Academy of Sciences for example (in Chinese). *Chin J Sci Tech Period*, 2019, 30: 870–877 [彭琳, 韩燕丽. 我国科技期刊数据政策分析及启示——以中国科学院主办英文期刊为例. *中国科技期刊研究*, 2019, 30: 870–877]
- 32 Shen Y. Data policy promotes integrative development of data publication and sci-tech journals in China (in Chinese). *Knowl Manag Forum*, 2021, 6: 252–262 [申艳. 我国科技期刊数据政策制定及实施模式研究. *知识管理论坛*, 2021, 6: 252–262]
- 33 Kong L H, Xi Y, Jiang L L. Open sharing and publishing policies for research data of scientific journals (in Chinese). *Chin J Sci Tech Period*, 2022, 16: 16–29 [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. *中国科技期刊研究*, 2022, 16: 16–29]
- 34 Niso G, Krol L R, Combrisson E, et al. Good scientific practice in EEG and MEG research: Progress and perspectives. *Neuroimage*, 2022, 257: 119056
- 35 Thompson P M, Andreassen O A, Arias-Vasquez A, et al. ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide. *Neuroimage*, 2017, 145: 389–408
- 36 Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI) – Establishing a PD biomarker cohort. *Ann Clin Transl Neurol*, 2018, 5: 1460–1477
- 37 Zhang G L, Wang J, Pan Y, et al. Scientific data sharing service mode and evolution (in Chinese). *Inf Stud Theory Appl*, 2022, 45: 70–77 [张贵兰, 王健, 潘尧, 等. 科学数据共享服务模式及其演化研究. *情报理论与实践*, 2022, 45: 70–77]
- 38 Wilkinson M D, Dumontier M, Aalbersberg I J, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 2016, 3: 1–9
- 39 Carroll S R, Garba I, Figueroa-Rodríguez O L, et al. The CARE principles for indigenous data governance. *Data Sci J*, 2020, 19: 43
- 40 Barker M, Chue Hong N P, Katz D S, et al. Introducing the FAIR Principles for research software. *Sci Data*, 2022, 9: 622
- 41 Shen Z H, Zheng X L, Zheng X H. PARIS principle: improving the usability of scientific data in the open collaborative environment. (in Chinese). *Big Data Res*, 2023, 9: 172–188 [沈志宏, 张晓林, 郑晓欢. PARIS原则: 开放协作环境下科学数据的可用性. *大数据*, 2023, 9: 172–188]
- 42 Hu L L, Zheng X H, Zhu Y H, et al. Research and practice on scientific data standard system in Chinese Academy of Sciences (in Chinese) *Library*, 2019, (11): 6–10 [胡良霖, 郑晓欢, 朱艳华, 等. 中国科学院科学数据标准体系研究与实践. *图书馆*, 2019, (11): 6–10]
- 43 Wang J L, Shi L, Xu B, et al. Study on the scientific data standards system (in Chinese). *China Sci Technol Res Rev*, 2020, 52: 45–51, 77 [王卷乐, 石蕾, 徐波, 等. 我国科学数据标准体系研究. *中国科技资源导刊*, 2020, 52: 45–51, 77]
- 44 Wang Z Q, Yang Q H. Research on the standards system of S&T resource management (in Chinese). *Stand Sci*, 2019, (6): 6–11 [王志强, 杨青海. 科技资源管理标准体系研究. *标准科学*, 2019, (6): 6–11]
- 45 Jiang T, Xu Z P, Chen X J, et al. Research on standardization and normalization of scientific data service (in Chinese). *China Sci Technol Resour Rev*, 2023, 55: 1–8, 93 [蒋甜, 许哲平, 陈学娟, 等. 科学数据服务标准化与规范化研究. *中国科技资源导刊*, 2023, 55: 1–8, 93]
- 46 Iso B S, Standard B. Information and documentation–The Dublin Core metadata element set. Dublin: ISO, 2009
- 47 Specka X, Gärtner P, Hoffmann C, et al. The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. *Comput Geosci*, 2019, 132: 33–41
- 48 Nichols T E, Das S, Eickhoff S B, et al. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat Neurosci*, 2017, 20: 299–303
- 49 Pernet C, Garrido M I, Gramfort A, et al. Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nat Neurosci*, 2020, 23: 1473–1483
- 50 Abrams M B, Bjaalie J G, Das S, et al. A standards organization for open and FAIR neuroscience: The international neuroinformatics coordinating facility. *Neuroinformatics*, 2022, 20: 25–36
- 51 Gorgolewski K J, Auer T, Calhoun V D, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*, 2016, 3: 1–9

- 52 Covitz S, Taper A M, Adebiyi A, et al. Curation of BIDS (CuBIDS): A workflow and software package for streamlining reproducible curation of large BIDS datasets. *NeuroImage*, 2022, 263: 119609
- 53 Corrado E M. Repositories, trust, and the CoreTrustSeal. *Technical Services Q*, 2019, 36: 61–72
- 54 Donaldson D R, Suleman H. Certification information on trustworthy digital repository websites: A content analysis. *PLoS One*, 2020, 15: e0242525
- 55 Audit and Certification of Trustworthy Digital Repositories. ISO 16363/CCSDS 652.0-M-1, <https://public.ccsds.org/Pubs/652x0m1.pdf> (2011)
- 56 Lin D, Crabtree J, Dillo I, et al. The TRUST principles for digital repositories. *Sci Data*, 2020, 7: 1–5
- 57 Hrynaskiewicz I, Simons N, Hussain A, et al. Developing a research data policy framework for all journals and publishers. *Data Sci J*, 2020, 19: 5
- 58 Markiewicz C J, Gorgolewski K J, Feingold F, et al. The OpenNeuro resource for sharing of neuroscience data. *eLife*, 2021, 10: e71774
- 59 Zuo X N, Anderson J S, Bellec P, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data*, 2014, 1: 1–13
- 60 Satterthwaite T D, Connolly J J, Ruparel K, et al. The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage*, 2016, 124: 1115–1119
- 61 Payakachat N, Tilford J M, Ungar W J. National database for autism research (NDAR): Big data opportunities for health services research and health technology assessment. *PharmacoEconomics*, 2016, 34: 127–138
- 62 Di Martino A, Yan C G, Li Q, et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry*, 2014, 19: 659–667
- 63 Alexander L M, Escalera J, Ai L, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data*, 2017, 4: 1–26
- 64 Taylor J R, Williams N, Cusack R, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 2017, 144: 262–269
- 65 Reer A, Wiebe A, Wang X, et al. FAIR human neuroscientific data sharing to advance AI driven research and applications: Legal frameworks and missing metadata standards. *Front Genet*, 2023, 14: 1086802
- 66 Paret C, Unverhau N, Feingold F, et al. Survey on open science practices in functional neuroimaging. *NeuroImage*, 2022, 257: 119306
- 67 Schwarz C G, Kremers W K, Wiste H J, et al. Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives. *NeuroImage*, 2021, 231: 117845
- 68 Jwa A S, Poldrack R A. The spectrum of data sharing policies in neuroimaging data repositories. *Hum Brain Mapp*, 2022, 43: 2707–2721
- 69 Wang J, Zhou X L, Shen Y Y, et al. Current status and prospects for scientific data's cross-border governance (in Chinese). *Chin Sci Bull*, 2024, 69: 1846–1856 [王珺, 周小林, 沈云怡, 等. 科学数据跨境流动治理的现状、思考与展望. *科学通报*, 2024, 69: 1846–1856]
- 70 Science D, Hahnel M, Smith G, et al. The state of open data 2023. *Digi Sci*, 2023. <https://doi.org/10.6084/m9.figshare.24428194.v1>
- 71 Chen R S, Zhou Y C, Jiang L L, et al. China Open Data White Paper 2023 (in Chinese). *Digi Sci*, 2023. <https://doi.org/10.6084/m9.figshare.24638301.v1> [陈润生, 周园春, 姜璐璐, 等. 中国开放数据白皮书2023. *Digi Sci*, 2023. <https://doi.org/10.6084/m9.figshare.24638301.v1>]
- 72 Tao R, Ye J Y. Scientific data citation: Sources, problems, and research trends (in Chinese). *Libr J*, 2022, 41: 19–29 [陶蕊, 叶继元. 科学数据引用: 源起、问题与研究趋向. *图书馆杂志*, 2022, 41: 19-29]
- 73 Altman Director o M, Borgman Professor C, Crosas Director o M, et al. An introduction to the joint principles for data citation. *Bull Asso Info Sci*, 2015, 41: 43–45
- 74 Cousijn H, Kenall A, Ganley E, et al. A data citation roadmap for scientific publishers. *Sci Data*, 2018, 5: 1
- 75 Fenner M, Crosas M, Grethe J S, et al. A data citation roadmap for scholarly data repositories. *Sci Data*, 2019, 6: 28
- 76 Madan C R. Scan once, analyse many: Using large open-access neuroimaging datasets to understand the brain. *Neuroinformatics*, 2022, 20: 109–137
- 77 Niso G, Botvinik-Nezer R, Appelhoff S, et al. Open and reproducible neuroimaging: From study inception to publication. *NeuroImage*, 2022, 263: 119623
- 78 Crüwell S, Apthorp D, Baker B J, et al. What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue

- of psychological science. *Psychol Sci*, 2023, 34: 512-522
- 79 Giehl K, Mutsaerts H J, Aarts K, et al. Sharing brain imaging data in the Open Science era: How and why? [Lancet Digital Health](#), 2024, 6: e526–e535
- 80 Li G J. AI4R: The fifth scientific research paradigm (in Chinese). *Bull Chin Acad Sci*, 2024, 39: 1–10 [李国杰. 智能化科研(AI4R): 第五科研范式. 中国科学院院刊, 2024, 39: 1–10]
- 81 E W N. AI helps to establish a new paradigm for scientific research (in Chinese). *Bull Chin Acad Sci*, 2024, 39: 10–16 [鄂维南. AI助力打造科学研究新范式. 中国科学院院刊, 2024, 39: 10–16]
- 82 Nature. Time to recognize authorship of open data. *Nature*, 2022, 604: 8

Summary for “人口神经科学数据共享: 政策生态、基础设施、实践和挑战”

Population neuroscience data sharing: Policy ecology, infrastructure, practices and challenges

Lulu Jiang¹, Peng Gao² & Yuanchun Zhou^{1,3*}

¹ Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China;

² The State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China;

³ Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

* Corresponding author, E-mail: zyc@cnic.cn

As a frontier in the field of brain science research, population neuroscience holds immense potential for realizing the translational application of neuroimaging to education, psychology, and medicine. However, its progress hinges on the presence of robust support for interdisciplinary data sharing. Motivated by the inherent need for disciplinary advancement, researchers have undertaken a multitude of bottom-up practices in data sharing to increase research reproducibility, reduce financial costs, and improve data reusability. These practices have ushered in the era of open science and initiated a profound transformation in research paradigms and academic exchange. In the contemporary landscape, global data sharing policies are becoming increasingly stringent, rendering data sharing an essential component of academic exchange. Nevertheless, the sustainable promotion of open data sharing still faces various impediments and challenges stemming from cultural, policy, and practical factors. Overcoming these obstacles requires collaboration among all stakeholders, including data contributors, consumers, communities, policymakers, and the general public. This article critically examines the indispensable elements required for the sustainable advancement of open scientific data sharing and delves into their interconnections, analysing the demands and contributions required from stakeholders in data sharing. It addresses scientific data sharing policies from the perspectives of diverse stakeholders, the norms and principles governing scientific data sharing, the standardized guidelines for data repositories, and the practical construction of data storage platforms specific to the domain of population neuroscience. Additionally, it supplements the discourse with examples from case studies such as theScienceDB - Chinese Color Nest Project Data Community. Progress in population neuroscience has benefited immensely from the paradigm of scientific data sharing, enabling researchers to leverage shared data sets to achieve innovative research and collaboration. However, this advancement has not been devoid of challenges. Privacy and security concerns, ethical considerations, the need for robust incentive mechanisms to promote data sharing, and deficiencies in scientific data management proficiency remain pressing issues in the current landscape of scientific data sharing. Nevertheless, with the advent of new scientific technologies, continuous improvements and safeguards for privacy and security are being realized. To address these challenges, this study proposes pertinent recommendations and prognostications concerning policy development, incentive mechanisms, infrastructure augmentation, and the trajectory of interdisciplinary convergence. These recommendations focus on the following key issues: (1) Strengthening top-level design to harness collective wisdom regarding innovation pertaining to data sharing forms and the exploration of issues of privacy, security, and standard specifications; (2) enhancing recognition of data sharing contributions and advancing incentive mechanisms; (3) leveraging high-quality data sharing services to improve standardization, data analysis, and visualization while providing accompanying personnel training to promote data sharing; (4) utilizing the advantages of platforms and AI models to create opportunities for interdisciplinary integration and cross-disciplinary innovation; and (5) exploring the open data journey in China by drawing insights from pathways in different cultural contexts.

scientific data sharing, data policy, data standard, data community, ScienceDB

doi: [10.1360/TB-2024-0147](https://doi.org/10.1360/TB-2024-0147)