arXiv:2603.19301v1 [cs.DL] 11 Mar 2026

材料科学期刊研究数据政策

Lukas Hörmann,[1,2] Hemanadhan Myneni,[3] Rwayda Kh. S. Al-Hamd,[4] Katarina Batalović,[5] Silvia Bonfanti,[6] Federico Grasselli,[7,8] Saulius Gražulis,[9] Bahattin Koç,[10] Konstantinos Konstantinou,[11] Ivor Lončarić,[12] Nataliya Lopanitsyna,[13] José Manuel Oliveira,[14,15] Paolo Pegolo,[16] Patrícia Ramos,[14,17] Kevin Rossi,[18,19] Sebastian P. Schwaminger,[20,21] Edith Simmen,[22] Milica Todorović,[11] Markus Stricker,[23,*] 和 Jonathan Schmidt[22,16]

[1]物理学院，维也纳大学，维也纳，*1090*，奥地利[2]化学系，华威大学，考文垂，*CV4 7AL*，英国[3]工业工程、机械工程与计算机科学学院，冰岛大学，雷克雅未克，冰岛[4]土木工程与管理系，理工学院，曼彻斯特大学，曼彻斯特，*M13 9PL*，英国[5]塞尔维亚共和国国家核科学研究所文查研究所，贝尔格莱德大学，塞尔维亚[6]*NOMATEN*卓越中心，国家核研究中心，*ul. A. Sołtana 7*，*05-400 Swierk/Otwock*，波兰。[7]物理、信息与数学科学系，摩德纳和雷焦艾米利亚大学，*41125* 摩德纳，意大利[8]*CNR-NANO S3*，*41125* 摩德纳，意大利[9]维尔纽斯大学，生命科学中心，生物技术研究所，*Saulėtekio al. 7*，*LT-10257* 维尔纽斯，立陶宛[10]萨班哲大学，*Orta Mah*，*FENS 1023*，*Tuzla*，伊斯坦布尔，土耳其[11]机械与材料工程系，图尔库大学，图尔库 *20014*，芬兰[12]鲁杰尔·博什科维奇研究所，*HR-10000* 萨格勒布，克罗地亚[13]先正达作物保护股份公司，*Schaffhauserstrasse*，*Stein*，*4332*，*AG*，瑞士[14]系统与计算机工程、技术与科学研究所，*4200-465* 波尔图，葡萄牙[15]经济与管理学院，波尔图大学，*4200-464* 波尔图，葡萄牙[16]计算科学与建模实验室，材料研究所，洛桑联邦理工学院，*1015* 洛桑，瑞士[17]*CEOS.PP*，*ISCAP*，波尔图理工学院，*4465-004* 圣马梅德德因费斯塔，葡萄牙[18]材料科学与工程系，代尔夫特理工大学，代尔夫特，荷兰[19]气候安全与保障中心，代尔夫特理工大学海牙校区，代尔夫特理工大学，*2594 AC*，海牙，荷兰[20]纳米实验室，药物化学部，奥托·勒维研究中心，格拉茨医科大学，*Neue Stiftingtalstraße 6*，*8010*，格拉茨，奥地利[21]格拉茨生物技术医学中心，*Mozartgasse 12*，*8010* 格拉茨，奥地利[22]材料系，苏黎世联邦理工学院，苏黎世，*CH-8093*，瑞士[23]先进材料模拟跨学科中心，波鸿鲁尔大学，*D-44799* 波鸿，德国（日期：2026年3月23日）

材料科学领域的开放且可重复研究依赖于数据、代码的可用性以及通用元数据标准。期刊研究数据政策（RDP）仍是制定和执行出版规范的主要机制。我们调查了涵盖17家出版商的171种材料科学期刊的研究数据政策，采用扩展的编码框架，既考察数据和代码共享行为，也考察同行评审标准。与早期关于研究数据政策的研究相比，我们发现了明显的进展迹象：几乎所有期刊都制定了研究数据政策，大多数期刊都提及数据可用性声明。然而，可执行的要求仍不常见，基础数据很少被要求必须公开存储，FAIR出版通常只是受到鼓励，而非强制要求。对研究软件的要求远不如对数据的要求成熟，对版本管理和持久标识符、依赖项披露、可重复执行环境或软件质量实践的关注都十分有限。将政策特征的研究结果汇总为开放研究数据评分后，我们发现各期刊之间存在显著差异。影响因子和访问模式都无法可靠地预测政策的严格程度。双重编码进一步表明，更复杂和更严格的政策可能更难以保持解读一致；我们还指出了不同研究在研究数据政策编码方面实现一致所面临的挑战。最后，我们提出了未来最佳实践的发展方向。

I. 引言

材料科学领域的研究因整合而取得了变革性的加速发展数据驱动方法和机器学习 [1–4]。随着研究人员依靠以数据为核心的工作流程，努力开发具有定制属性、适用于储能、能量转换和电子学等领域的创新材料，数据可用性的需求比以往任何时候都更加迫切。要获得可验证的结果，就必须系统地管理和共享底层

* 通讯作者：markus.stricker@rub.de

研究成果，包括所有数据集，以及用于生成和分析这些数据集的软件和工作流程；仅靠稿件本身的报告远远不够 [5–7]。数据共享、代码共享和竞争性基准测试的环境，理想情况下能够促成一种无障碍可重复性，这一直是机器学习 [8] 及其在材料科学中的应用取得进展的关键驱动力。

践行开放研究实践的承诺能够促进合作、加快发现、验证结果，并推动工具在不同学科间的重用 [3, 9]。尤其对于数据驱动方法而言，高质量、机器可读数据集的可用性是构建稳健模型的必要基础 [10]。

不同科学学科的数据共享实践差异显著。在晶体学和结构生物学领域，数据共享已相当成熟：20 世纪 70 年代成立的蛋白质数据库（PDB） [11]逐渐发展为全球大分子结构档案库 [12, 13]。到 2008 年，不仅结构坐标，结构因子和核磁共振约束也必须提交 [14], ，从而能够验证原始研究结果，并在某些情况下加以改进。这项由全球 PDB 联盟协调、主要出版商和资助机构支持的政策 [15] ，显著提高了结构数据的可用性，并促成了机器学习的成功案例，例如蛋白质折叠预测 [16–19]。近来，计算和存储技术的进步使原始衍射数据和显微镜数据的代码仓库成为可能 [20]。

在材料科学社群中，数据共享同样被视为推动该领域发展的关键，这一点从采用 FAIR（可查找、可访问、可互操作、可重复使用）等国际框架原则便可见一斑，这些原则适用于材料科学中的数据 [7] 和代码 [21] [22]。虽然目前尚无专门存储晶体结构数据的单一集中式仓库，但主要计算数据库 [23–28]的维护者已整合 OPTIMADE API 标准 [29], ，实际上为基础的模拟晶体结构数据创建了一个通用数据空间。尽管晶体结构数据只是多元材料科学研究数据中的一个小型子空间，这些成功的数据共享倡议表明，可以调整最佳实践，以弥补材料数据共享标准方面的不足。

多个正在进行的项目和社群致力于制定数据和元数据标准，例如 NOMAD [26], 、电池数据基因组 [30], 、材料基因组计划 [31], 以及材料研究数据联盟 [32], ；同时，人们也提出了各种策略，鼓励作者在各个层面共享数据。这包括资助机构 [33–38], 、大学 [39, 40], 、研究社群 [41], 和个别研究团队发起的倡议。

尽管全球社区付出了努力，但数据格式缺乏标准化、不愿分享专有或敏感数据、对分享数据和代码缺乏信心等问题，以及存储和访问基础设施不足，仍持续阻碍着进展 [30, 42, 43]。此外，材料科学数据复杂、异质且数量庞大，其中许多数据是在缺乏成熟数据整理框架的情况下生成的，这些特点持续为有效的数据管理带来障碍 [44]。再加上机器学习驱动的研究迅速扩张，这些局限性又带来了第二种更具系统性的风险：质量低、文档记录不完善或验证不足的数据集和工作流程可能会被大规模重复使用，在后续研究以及广泛使用的数据库中悄无声息地传播错误，产生噪声和错误结论。因此，加强数据管理维护和可复现性实践，不仅关乎提高数据管理效率，也是维护可信度的先决条件；随着机器学习（ML）日益深入地融入材料科学，这一点尤为重要。

科学、技术、工程和数学学科（科学、技术、工程和数学）领域的研究表明，期刊是社区实践的仲裁者，甚至比资助机构更能激励作者遵循社区特定的最佳实践，推动该领域发展 [45] 。然而，人们常发现，期刊特定政策的标准不够完善，有些情况下甚至完全缺失 [46–49]。例如，Rousi 等人（2020） [47, 50报告称，他们考察的高被引物理学期刊中有 35% 没有研究数据政策（RDP）；而 Resnik 等人（2019）62]发现，43.6% 的受考察期刊在其政策中完全没有涉及数据共享。

要采用更全面、更严格的研究数据政策，面临着系统性挑战。首先，这类政策不可避免地会增加研究人员、审稿人和编辑的工作量 [63], ，这可能与营利性期刊的激励机制相冲突。更严格的标准可能会拖慢出版流程，减少接收的稿件数量，进而降低收入；在收入与发表量挂钩的开放获取模式下，这种影响尤其明显。其次，在许多材料科学领域，公司是研究的重要贡献者，但往往无法共享数据 [64]。严格执行数据发布标准可能会减少它们对公共研究的科学贡献。另一方面，审稿人和编辑需要获取研究结果和代码，才能对研究的可重复性作出有依据的判断。目前，在稿件接收之前，坚持良好实践在很大程度上取决于个别编辑和审稿人；但如果没有明确的期刊数据和代码共享政策，他们的权限有限。值得肯定的是，目前已有旨在自动化落实良好代码共享实践的举措，例如 FACILE-RS [65]。

然而，其他领域的纵向研究表明，研究数据政策的适应工作取得了进展：在数学和多学科科学期刊中，没有数据政策的期刊占比从2011年的66%降至2012 [46], 年的62%；而在社会科学领域，这一比例则从

2003年的90%降至2015年的61%。文献 [51]对这些研究及相关研究进行了全面综述。

在本文中，我们扩展了现有的期刊研究数据政策编码框架 [50, 51] ，以定量评估材料科学领域当前的政策格局。这里所说的“编码”遵循社会科学中的含义，即系统地对定性和定量信息进行分类，以便开展结构化分析。我们调整后的框架纳入了材料科学特有的考量，明确评估了超越代码共享的研究软件（包括对代码标准和计算工作流可复现性的要求），并捕捉了编辑和审稿人指导等执行机制（完整框架见第五节）。我们使用 Dimensions [66] 应用程序接口和选定的研究领域定义，整理出一个包含来自 17 家出版商的 171种期刊的数据集，该数据集广泛代表了材料科学的发表渠道。对于每家出版商，我们抽样选取了——在可获取的情况下——影响力最高的三种期刊（按期刊影响因子计算）、发表量最大的三种期刊，以及随机选取的三种期刊，以检验影响力或受欢迎程度等期刊特征是否与政策严格程度相关。此外，我们还纳入了少量以数据为重点的期刊（例如，*CODATA*《数据科学期刊》），以对标面向最佳实践的政策。最后，我们考察了对研究数据政策进行一致编码所面临的挑战，包括在单项研究内部保持一致，以及比较不同研究之间的结果。

II. 结果

图1概述了有关研究数据政策基本要素的结果。这些结果包括：a）是否存在研究数据政策；b）数据可用性声明（DAS）相关规定；c）数据共享要求的具体内容；d）代码共享要求的具体内容；e）数据发布时机；以及 f）期刊的审稿人指南是否提及数据和/或代码。

数据政策。 首先，研究数据政策呈现积极趋势：与早期研究相比，我们调查的期刊中有 [46, 50, 51], 98% 制定了研究数据政策（图1 a），95 % 提及数据可用性声明（图1 b）。然而，近半数期刊仍未要求提供数据可用性声明。尽管众所周知，仅有数据可用性声明不足以建立强有力的数据出版文化 [67, 68], ，但我们仍将其视为研究数据政策的基础之一，前提是在发表前的审稿流程中核实数据可用性声明的真实性。

大多数研究数据政策都提到了数据共享要求，但通常将其列为“可选”（图1 c）：只有 7 % 的期刊在所有情况下都要求公开共享所有研究数据。针对特定数据类型（例如晶体结构、基因组学）的要求较为常见，适用于 11 % 的期刊。另有 1.8 % 的期刊要求在同行评审期间共享所有数据，1.2 % 的期刊要求在提出合理请求时共享时共享。

鉴于先前研究发现，仅有数据可用性声明不足以促进良好的数据共享行为 [67,68], ，我们还调查了数据发布时机（参见图1 e）以及使数据公开可用的要求。Rousi 等人 [51] 发现，在调查的40种物理学期刊中，有15种要求在发表时发布数据。在我们的材料科学样本中，将“要求在审稿流程之前提供数据”和“要求在正式发表之后提供数据”这两类合并后，我们发现，36.3%的期刊明确要求在审稿和/或发表时提供数据可用性。这一比例总体上与文献 [51]报告的相应比例相当，但略低于后者。

到目前为止，我们只讨论了研究数据是否以及何时被共享。共享数据的位置同样重要，因为这会影响数据的可发现性、长期保存以及对元数据标准的遵循。具体而言，三分之二的期刊鼓励在 FAIR 仓储上共享数据，11.9 % 的期刊要求在 FAIR 仓储上共享某些数据类型，而只有1.2 % 的期刊要求将所有数据共享至 FAIR 仓储。相反，77.4 % 的期刊在其开放数据政策中未明确提及数据许可。根据 FAIR 定义 [7], ，“可重用”被定义为“（元）数据以清晰且易于获取的数据使用许可发布”。无论如何，许可的存在本身并不意味着该许可是允许使用的许可，也不意味着数据本身具有可重用性。因此，旨在促进数据共享的期刊政策应支持允许使用的许可。

严格的政策如果缺乏执行，裨益甚少，而执行通常需要审稿人的专业知识。然而，很少有期刊考虑数据在同行评审流程中的作用，61%的期刊在审稿人指南中没有提及研究数据政策（图1 f）。提及该政策的期刊不到三分之一，只有极少数期刊在审稿人指南中明确要求共享数据和/或代码，或说明会指定一名数据审稿人。

代码政策。 与数据政策形成鲜明对比的是，在所有受调查的期刊中，软件代码政策都较为宽松。70%的期刊鼓励但不强制代码共享，仅有1.8%的期刊要求代码共享。即使在鼓励代码共享的期刊中，超过80%的期刊也不要求作者注明软件依赖项或执行环境。超过80%的政策未就代码的持久标识符提供指导。提及编码标准（例如代码风格或安全结构）的情况并不常见——约80%的政策对此未作规定；要求任何最低标准的期刊不到10%，而提及测试和代码检查的情况则更为罕见。

开放数据评分

为总结政策严格程度，我们报告了开放数据评分 （开放数据评分； $s \in [0,1]$， = 越高表示政策越严格；完整定义见第 VC 节）。我们评估期刊的平均开放数据评分，

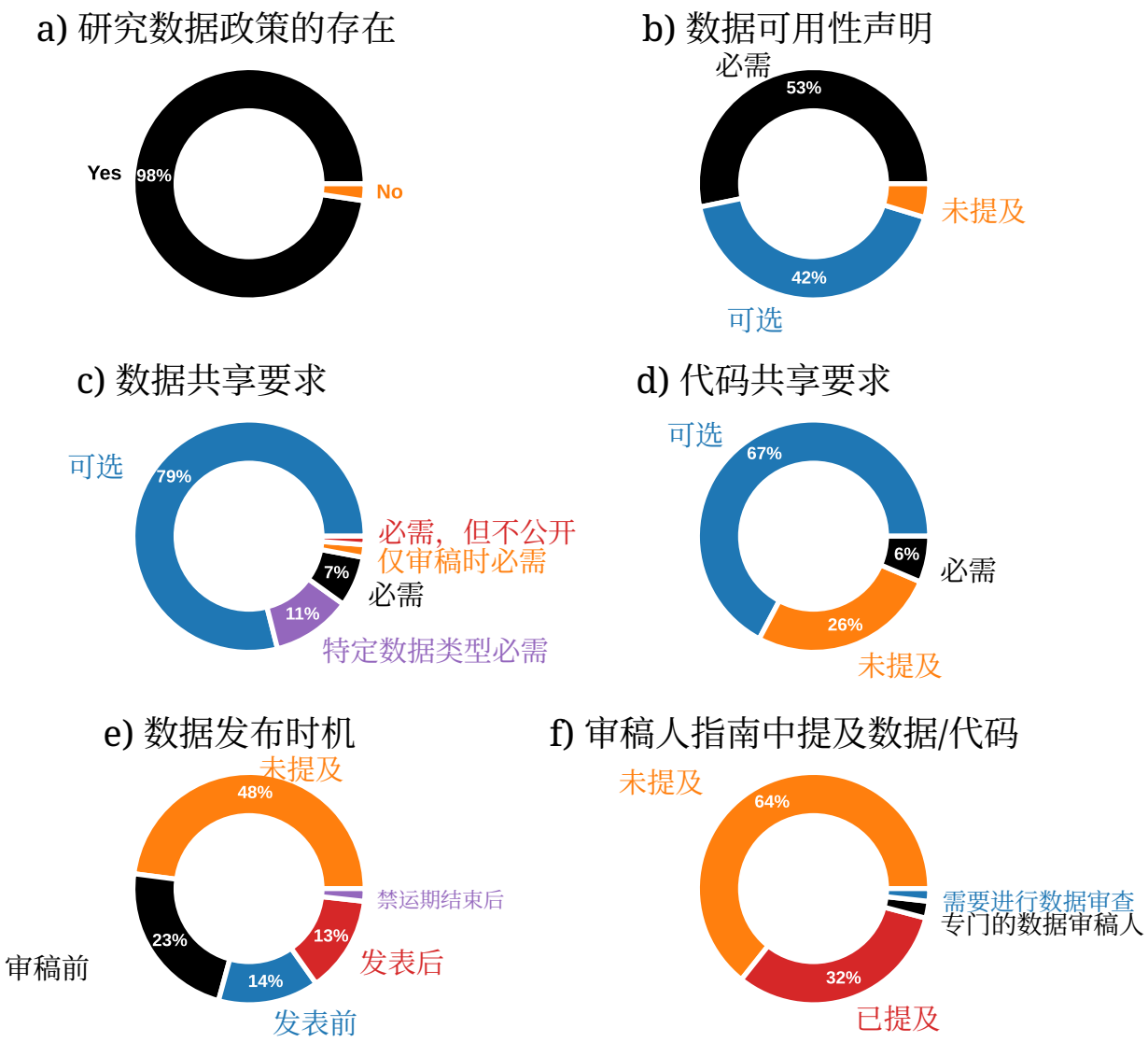


图 1：数据和代码共享要求概览。

而每个问题在各期刊间的差异则通过逐题标准差来衡量。这些结果见图2。

期刊在是否制定研究数据政策、是否要求提供数据可用性声明，以及数据共享方式政策方面的开放数据评分（ODS）相对较高；许多期刊还推荐使用 FAIR 数据存储库，并重视数据可引用性。相较之下，对数据共享的具体要求所获开放数据评分较低。例如，数据共享要求的开放数据评分低于 0。2。总体而言，我们发现标准差较大。例如，数据可用性和数据发布时间的标准差较大，表明各期刊在是否要求提供相应的数据可用性声明，以及必须何时发布数据方面存在显著差异。

我们检验了若干关于严格程度较高的研究数据政策相关因素的假设，重点关注开放数据评分与以下因素之间的关系：（i）期刊影响因子，（ii）学会出版商与非学会出版商，（iii）开放获取与非开放获取期刊，以及（iv）期刊创刊年份。

图3a展示了开放数据评分与期刊影响因子的关系。尽管一些高影响因子的期刊开放数据评分较高，但两者之间没有明显的单调趋势：影响因子范围内的各类期刊，其开放数据评分覆盖了全部取值范围，这表明声望（影响因子）与政策严格程度并不存在可靠的相关性。皮尔逊相关系数为$r = 0.175$ ($n = 169$, $p = 2.3\%$)，表明两者的关联较弱，但在统计学上仍具有显著性。为评估极端值的影响，我们在排除影响因子最高的两本期刊后重复了分析（预先设定的规则：影响因子 > 75）。相关性消失了（$r = -0.007$, $p = 93.9\%$），这表明表面上的关联是由这些离群值而非普遍趋势所驱动的。

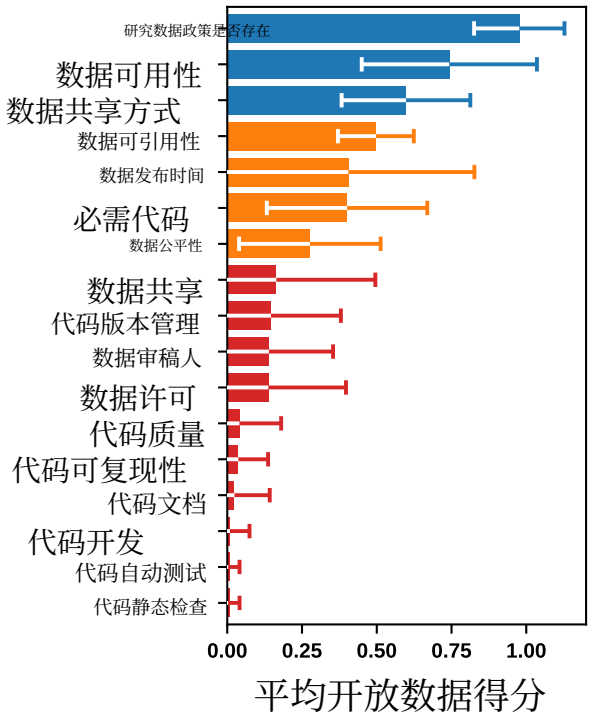


图2：各问题的平均开放数据得分；误差线表示开放数据评分的标准差。颜色表示开放数据评分低于0.2（红色）、低于0.5（橙色）和高于0.5（蓝色）的问题。

总体趋势。

图3b展示了开放数据评分与成立年份之间的关系。皮尔逊相关系数为 $r = 0.166 (n = 171, p = 3.0\%)$，表明两者之间存在较弱但在统计学上显著的关联。

图3c展示了出版商类型（学会或营利性）与开放数据评分之间的关系。尽管我们发现营利性出版商的平均开放数据评分高于学会出版商，但方差分析F检验和t检验均未发现具有统计学显著性的相关性。

图3d展示了期刊的开放获取政策（完全开放获取或混合型）与开放数据评分之间的关系。我们发现，开放获取期刊的平均开放数据评分高于采用混合型政策的期刊。为评估其显著性，我们进行了方差分析F检验和t检验。F统计量为19.702，p值为0.002 %，表明这一结果具有统计学显著性。

编码过程的一致性

如第五节B所述，两名研究人员独立对每项研究数据政策进行编码，随后比较结果以确保一致性。典型的不一致情况出现在同一政策的不同章节使用了相互矛盾的措辞时。例如，在一处要求将数据存入 FAIR 存储库，而在另一处仅鼓励这样做。根据我们的保守规则，此类情况编码为鼓励。对于在所有期刊中采用单一、简短且统一政策的出版商（例如编码时的 APS），研究人员之间的一致性最高；而对于更复杂、篇幅较长且附有针对特定期刊附录的框架（例如 Frontiers、MDPI、RSC），则出现了更多分歧。

图4a概述了编码过程。75%的问题最初编码一致，而最初编码不一致的问题可归因于遗漏了RDP文本（占所有问题的13%）以及对政策含义的误解（12%）。

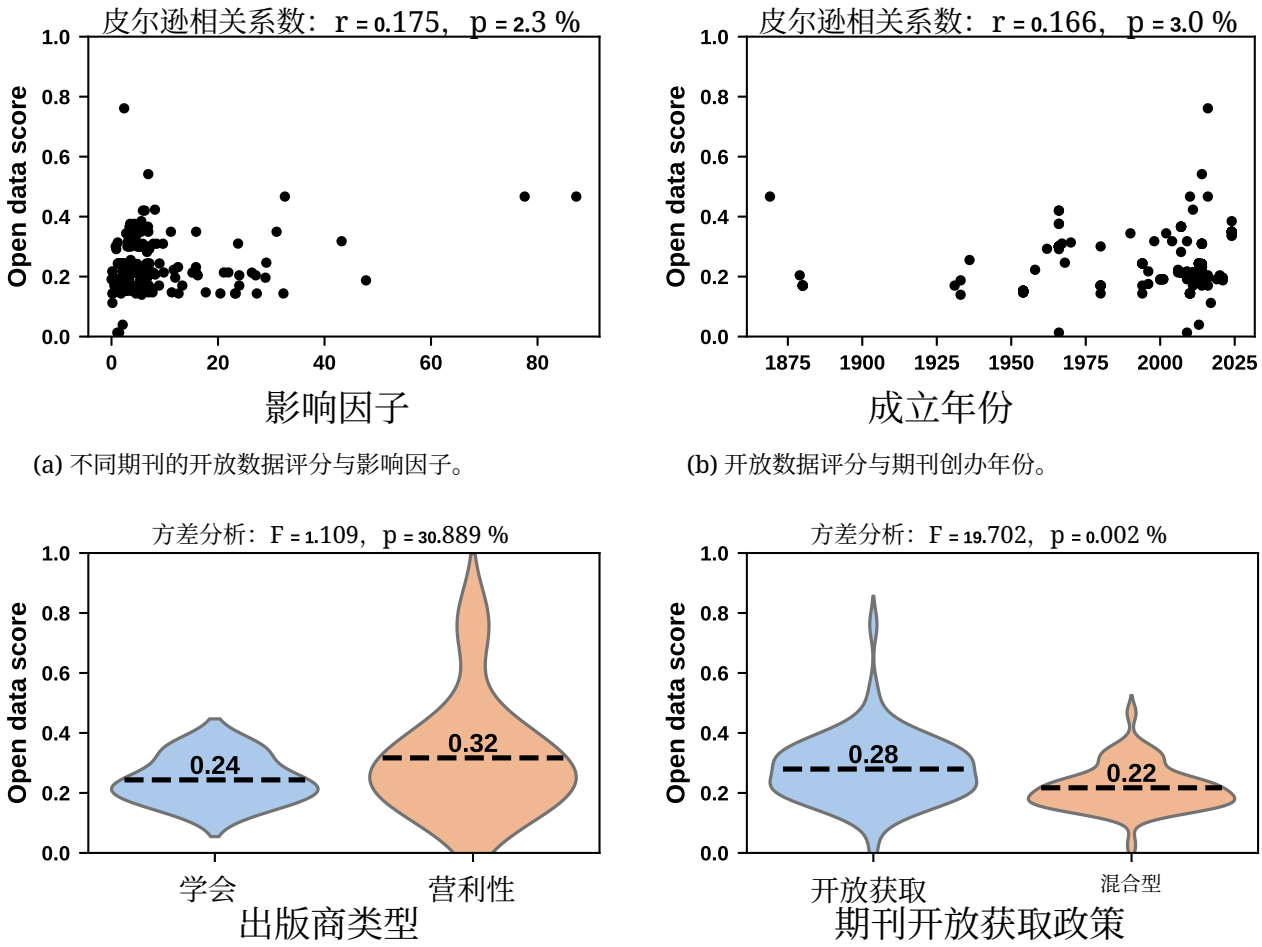
编码最不一致的条目是问题10和11（建议/要求采用专属政策的数据类型）、问题7（数据共享方式）和问题6（数据发布时机）。问题10和11的不一致程度较高是意料之中的，因为这两题需要通过自由文本提取特定数据类型的规则。对于问题7，选项“*RDP*中同等推荐的多种数据共享方式”占其不一致情况的75%，这表明该问题表述不当，应在未来研究中修订。对于问题6，政策文本中“要求在发表前”与“要求在发表后”之间的区别往往不明确，这表明期刊可以更明确地说明发布时间安排。

这些结果促使我们事后提出一个假设：政策严格程度会影响编码一致性（图4b）。我们将开放数据评分作为政策全面程度的替代指标，并从一致性评分中排除问题10和11，观察到皮尔逊相关系数为 $r = -0.615$、$p = 0.9\%$ 的强负相关：政策越全面，往往越难以保持一致编码。由于这一假设是在分析阶段提出的，因此应将这些发现视为用于假设生成的结果，并且需要在未来的研究中加以验证。

III. 讨论

我们的结果表明，材料科学期刊在很大程度上已就开放研究的表述方式（例如政策和数据可用性声明）达成一致，但尚未形成共同且可执行的基准，无法可靠地产生可重复使用的数据集和软件。其结果是政策格局支离破碎，实际的解读和合规负担被转嫁给作者、编辑和审稿人。实际上，这意味着已发表研究的可重复性仍可能取决于出版平台特定规范，而非社区共同期望。

数据与代码政策。 材料科学期刊中，有相当一部分（36.3 %）要求在发表时提供数据可用性，这一比例略高于此前针对物理学期刊的报告 [51]。然而，大多数期刊仍允许延迟或有条件地发布数据，这削弱了即时复现的目标。或许最令人担忧的是以下关于



(a) 不同期刊的开放数据评分与影响因子。

(b) 开放数据评分与期刊创办年份。

(c) 开放数据评分与出版商类型；每种类型的平均开放数据评分出版商被视为一个样本。

(d) 期刊的开放数据评分与开放获取政策。

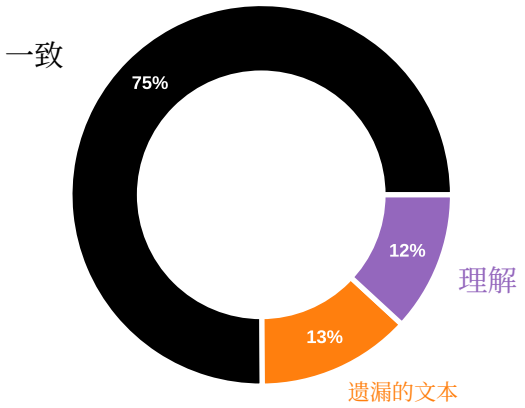
图3：开放数据评分与期刊特征。

数据共享 本身的结果。要求公开存储所有数据的期刊不到 7 %，而大多数期刊仅仅鼓励共享数据。政策措辞中理想化的目标与可执行要求之间的这种脱节，可能导致期刊政策与作者实践之间的差距长期存在。

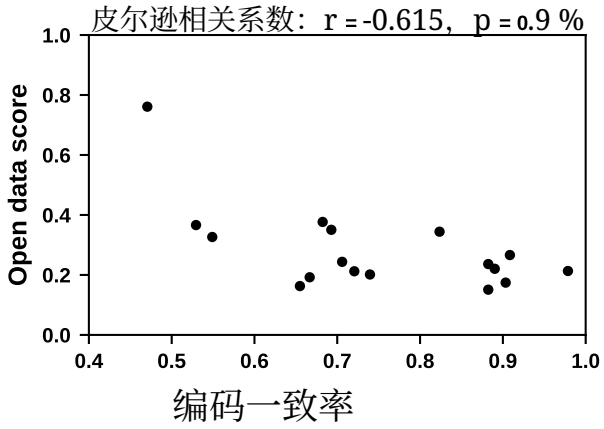
数据共享方式是一个几乎无需增加成本便可改进的领域。鼓励在 FAIR 仓储上共享数据不需要额外资源；对于已有仓储和标准的数据类型，没有相应政策的期刊也可以轻松采用现有建议。我们在这方面观察到了积极趋势。期刊经常鼓励使用 FAIR 仓储，也常提及可引用性，这反映出人们对数据管理最佳实践已有认识。不过，对符合 FAIR 标准的出版物提出要求的情况仍然少见，而大多数政策（77.4 %）缺乏许可指导，这令人对共享数据的可复用性产生疑问。

一项政策的实际效果取决于它是否融入编辑工作流程。尽管审稿人指南中提及数据/代码的情况比此前报道的更为常见 [50], ，但大多数期刊仍未将数据和软件视为同行评审期间常规检查的对象。评估某个数据集是否确实能使文章具备可重复性并非易事；除非在同行评审流程中对此进行核实，否则之后很难确认这一点。在我们的研究中，我们仅发现三种期刊设有数据和/或代码审稿人，负责检查数据是否完整，并确认代码的基本功能。

编码标准和质量保证是工业软件可靠性的基石，尤其是在汽车行业 [69]或航空电子 [70]等高责任领域。显然，科学软件开发者仍有很多东西要向工业软件开发者学习 [71]。与先前研究相比，Stodden 等人 [46] 发现，计算与数学生物学期刊中有 22 % 制定了代码政策，类似于《参考文献》 [71]中物理科学期刊的 19.1 %，



(a) 编码问题所占百分比汇总
由每本期刊的两位编码者一致作答
以及不一致的原因。



（b）开放数据评分与编码一致性之间的关系，不包括问题10和11。

图4：编码一致性及其与开放数据评分的关系。

因此，70 % 可视为一项进步。然而，Resnik等人 [50] 还发现，20.1 % 的期刊要求将代码存入代码仓库，这与我们的数据集中 6.4 % 的比例形成鲜明对比。我们分析了《参考文献》 [50]中被认定要求代码存 deposited 的 17 种物理学期刊的研究政策，发现其中 11 种期刊并无代码存 deposited 要求。我们认为这种差异可能有两种解释。首先，有关代码的研究数据政策可能有所退步。鉴于我们的研究及其他研究 [51]报告的开放数据要求总体上有所加强，这种情况似乎不太可能。一个更合理的解释是，在不同研究之间及单项研究内部，始终如一地对研究数据政策进行编码存在挑战。例如，我们的研究对“要求”一词采取了严格解读，如第五节所述，而其他研究可能采用更灵活的定义。此类解读差异很容易导致结果不一致。在此背景下，我们建议未来研究最好将研究数据政策中的具体文本摘录作为数据的一部分纳入其中，以便说明其编码决策，从而更好地回答未来此类问题。此外，我们发现，即使在我们自己的编码过程中，不同编码者得出的结果也有所不同，详情见第五节B和补充信息中的第?? 节。这种差异可能不仅反映了编码框架的不足，也体现了某些开放数据政策本身措辞的含糊性。

研究数据政策中几乎从未提及代码质量，这一点从代码检查和自动化测试的研究数据得分几乎为零便可看出。对代码质量标准缺乏重视，部分原因可能在于大多数材料科学研究人员并未接受过正规的软件开发培训，而且代码质量通常也不是研究的重点。由于良好的软件实践需要投入大量时间和资源，资助机构在材料科学领域很少将其列为优先事项。鉴于研究代码具有广泛的多样性——从简短的分析脚本到大规模的模拟软件包——在期刊层面很难强制执行严格且普适的标准。不过，对于高度重视方法和软件开发的期刊而言，要求更高的代码质量标准显然是有必要的。

最后，开放数据评分与编码一致性之间的负相关（图 4b）凸显了政策设计中的一种权衡：试图涵盖许多情况和数据类型的政策，可能会变得更难以解读和实施。这表明，加强研究数据政策不仅仅是增加更多要求，还要使要求可检验且明确无歧义。因此，一种切实可行的方向是，使政策与现有标准和基础设施保持一致，同时将其表述为一小组检查项，以便在编辑和同行评审流程中进行核验。具体而言，未来的政策模板可以优先考虑：（i）简明、措辞不冲突的核心内容；（ii）明确分级的义务（例如，必须 → 预期（可能影响编辑决定） →提及）；以及（iii）针对时间安排、位置（FAIR 仓储）、持久标识符和许可的规范决策路径，并附上期刊专属附录；除非明确规定优先级规则，否则这些附录不得推翻核心内容。

我们的结果指出，少数反复出现的缺口已有成熟的社区标准充分涵盖，却很少在期刊政策中成为可执行的要求。不过，本研究在选择哪些期刊进行研究数据政策评估方面存在局限。下文将这些缺口对应到期刊可以采纳的政策条款和检查项，而无需进行学科特定的重新发明。

tion.

从观察到的差距到政策模板

（1）从“鼓励”转向可验证的存档。虽然许多期刊鼓励将数据存档至存储库，但只有少数期刊要求进行符合 FAIR 的存档，而且大多数政策都缺少许可指导（77。4%）。因此，可执行的最低基准应包括：(i) 将重现主要论断所需的所有数据存档至社区认可的存储库（如有条件，则符合 FAIR），(ii) 提供持久标识符（通常为 DOI），以及 (iii) 明确的再利用许可证。这些要素均有成熟的基础设施支持，例如 DataCite 和 DOI [72, 73]。编辑部的核验可以从简：数据可用性声明必须包含可解析的标识符和许可证，且审稿人必须在审稿时确认已存档材料可供使用。

(2) 将研究软件视为首要研究成果。我们的调查显示，代码要求远远落后于数据要求：代码共享通常是可选的，而且大多数政策都没有涉及依赖项、版本和可复现的执行环境。与之对应的基线要求应与 FAIR4RS 保持一致 [74], ：归档并发布带版本的代码，并为其分配持久标识符，同时明确说明执行环境。例如，requirements.txt/environment.yml 是适用于 Python 代码的最低限度示例，但无法实现完全可重复性；而较新的环境管理器 [75]或容器配方则能够实现完全可重复性。在可行的情况下，期刊可以要求提供一个统一的“复现”入口，用于重新生成关键图表或表格。现有代码仓库和归档服务（例如 Zenodo [76]）、常用环境管理器（例如 Conda/Spack/Pixi）以及容器化方法都为此提供了支持。重要的是，这些要求可以根据适用范围进行调整：与主要采用实验方法的研究相比，期刊可以对方法/软件论文提出更严格的要求。

（3）代码质量：关注最低限度的信号，而非全面的工业级合规。由于科学软件的形态从简短脚本到大型模拟软件包不一而足，在期刊层面要求普遍适用的行业级标准并不现实。不过，期刊仍可要求一些最低限度的质量信号，以提升可维护性和可复用性：对输入/输出进行基本文档编制、提供许可证，以及在软件对研究贡献至关重要时，至少对核心程序进行一项自动化合理性检查（单元测试或功能测试）。这类要求符合既有的科研软件工程指导原则 [77, 78]，也有助于应对当前政策近乎缺失的状况。

鉴于大多数期刊仍未将研究数据政策要求纳入审稿人指南，一个关键步骤在于程序层面：要求投稿时附上简短的可复现性清单，内容包括数据 DOI + 许可、代码发布 + 版本、环境规范，以及图表与相关产物的对应关系。在工作量允许的情况下，期刊可以借鉴少数期刊中正在形成的实践，为部分投稿配备专门的数据/代码编辑。这一做法针对的是政策措辞与实际可验证性之间已观察到的差距，同时不会将全部负担转嫁给单个审稿人。

（6）超越逐个存储库合规：将互操作性纳入未来政策维度。最后，我们的调查表明，当前研究数据政策很少涉及跨存储库和跨学科的互操作性。研究数据联盟和 FDO 论坛 [79–82], 讨论的 FAIR 数字对象（FDO）等新兴概念，旨在支持标识符的机器可操作解析，以及跨领域访问异构数据。尽管 FDO 系统仍在发展中，但政策中完全缺乏有关跨领域互操作性的表述，这表明，面向未来的研究数据政策模板有机会提前应对材料数据在各领域间日益增加的复用需求。

(7) 将能力建设纳入政策设计。最后，只有研究人员能够切实遵守更严格、更关注软件的政策，这些政策才会有效。由于许多材料科学家接受的数据管理和软件工程正规培训有限，期刊和出版商可以明确向作者推荐社区培训资源（例如 TheCarpentries 和 CodeRefinery），以减少遵循政策的阻力；这些资源涵盖版本控制、许可、文档编制、测试、持续集成和可复现环境 [83, 84]。在研究数据政策中引用此类资源，既能在执行要求的同时提供支持，也能直接弥补可能导致代码政策采纳率偏低的技能差距。

IV. 结论

我们拓展了现有编码框架，使其涵盖数据和代码共享要求，并将其用于对171种期刊的材料科学研究数据政策进行定量调查。总体而言，政策的制定和意向方面取得了明显进展，但可执行的要求仍然少见，研究软件政策的滞后则更为显著。将政策特征汇总为开放数据评分后，我们发现各期刊之间存在显著差异，而期刊影响因子和访问模式对政策强度的预测能力有限。最后，双重编码表明，政策的复杂性和严格程度可能降低其可理解性，这凸显了采用简洁、内部一致的表述以及明确分级的义务要求的必要性。

这些发现共同为材料科学领域当前研究数据政策（RDP）的实践提供了基准，也为制定以证据为依据的政策模板提供了依据，将社群标准转化为一小组明确无歧义、可验证的

检查项，以便在投稿和同行评审期间进行核查，同时兼顾未来需求，例如跨存储库互操作性。

V. 方法

A. 期刊选择

为了选出一组具有代表性的期刊，我们使用了 Dimensions [66]Python API。首先，我们将选择范围限定在澳大利亚和新西兰标准研究分类（ANZSRC 2020）定义的 9 个研究领域：

- 3047 理论与计算化学
- 3403 大分子与材料化学
- 3402无机化学
- 3405有机化学
- 3406物理化学
- 4016材料工程
- 4018纳米技术
- 5102原子、分子与光物理学
- 5104 凝聚态物理学

按照文章数量搜寻最受欢迎的出版商后，我们初步筛选出一批出版商。对于每家出版商，我们随后选取引用次数平均值最高的三种期刊、最受欢迎的三种期刊，以及三种随机期刊。对于期刊数量少于九种的出版商，则选取其所有可用期刊。用于获取数据的详细查询以及由此生成的数据集，均随文章一同发布。由于编码期刊耗时较长，我们将范围限制在最受欢迎的17家出版商，共计171种期刊。

B. 编码过程

两名研究人员分别对每家出版商的研究数据政策进行编码。对于每个答案，编码人员都会保存支持其选择的确切政策文本（若不存在相关政策文本，该字段必然为空）。独立编码完成后，其中一名编码人员会比较两组结果，以确保一致性。如有差异，审阅者会尽可能根据收集到的文本选出正确答案；若无法做到，则与研究团队其他成员讨论确定正确答案。我们还会记录并分析差异产生的原因，以改进编码框架和未来的研究数据政策。如果存在多种可能的解读，或不同文本部分提供了相互矛盾的信息，则采用最保守的解读。编码过程于2024年6月至2025年5月进行。对于每本期刊，已发布数据中均注明了编码日期。我们注意到，研究期间研究数据政策已经有所改进，例如，APS 的研究政策变得明显更加严格。我们希望几年后能对这一进展进行系统比较。

C. 数据政策的数值评分

开放数据评分 依赖于这样一个事实：我们的编码框架具有排他性，并按层级排序。每个问题 n 的每个答案 α 都会被赋予一个严格程度值$w_{n\alpha}$ ，范围为0（最宽松）到1（最严格）。我们用 α_n表示每个问题所选的答案。要计算开放数据评分 （s），我们将所有问题的得分相加，并除以问题数量 N进行归一化。

$$s = \frac{1}{N}\sum_{n=1}^{N}\sum_{\alpha=1} w_{n\alpha}\delta_{\alpha,\alpha_n} \tag{1}$$

此处，$\delta_{\alpha,\ \alpha_n}$ 是克罗内克δ函数，用于指示是否选中了给定答案。

D. 编码框架

为了客观且可重复地评估开放数据政策，我们制定了一个编码框架，其形式为互斥的选择题。各个选项按照其体现开放数据原则的程度排序，越靠后的答案代表越有力的政策。这使得我们能够将其关联起来为每个问题赋予一个数值。我们的编码框架在先前工作的基础上构建 [50, 51], ，强调清晰的可排序性，并纳入其他相关要素——例如科学代码。我们将与文献{v1}和{v2}重叠的问题分别与参考文献 [50]和 [51] 重叠的引用。编码框架定义如下：

1. 研究数据政策的存在情况 [51]

（a）否

（b）是

2. 数据共享要求（改编自 [50]）

（a）仅与编辑和审稿人共享数据（不公开共享）

（b）鼓励但不强制数据共享 （c）要求共享数据但不公开（例如，可应要求提供） （d）要求对特定数据类型进行公开数据共享 （e）要求公开共享所有数据

3. FAIR 数据共享

(a) 未提及在 FAIR 仓储中公开共享数据 (b) 鼓励在 FAIR 仓储中公开共享数据 (c) 要求特定数据类型在 FAIR 仓储中公开共享 (d) 要求在 FAIR 仓储中公开共享所有数据

4. 数据可用性声明

(a) 未提及

(b) 已提及但可选 (c) 必需

5. 数据的可引用性和可查找性

(a) 未提及 DOI 或其他持久标识符（Handle、档案资源密钥（ARK）等）(b) 建议为数据集或代码提供 DOI 或其他持久标识符 (c) 要求为数据集或代码提供 DOI 或其他持久标识符

6. 数据发布时机[51]

(a) 未说明数据何时可用 (b) 要求的数据须在审稿前提供 (c) 要求的数据须在发表前提供 (d) 要求的数据须在发表后提供 (e) 要求的数据须在禁运期结束后提供

7. 推荐的数据共享方法（改编自 [50]）

(a) 未推荐任何方法 (b) 同等推荐多种方法 (c) 应作者要求共享数据 (d) 推荐使用补充材料或期刊托管共享 (e) 推荐使用公共在线存储库 (f) 推荐使用 FAIR 仓储

8. 建议/要求采用的许可证

(a) 未提及许可证 (b) 提及特定许可证类型 (c) 要求共享的数据采用开放获取许可证 (d) 要求采用开源许可证

9. 关于研究数据的审稿人指南

(a) 审稿指南中未提及数据共享政策 (b) 审稿指南中提及数据共享政策 (c) 审稿人确认共享的数据/代码 (d) 需要额外的数据/代码专门审稿人

10. 列出建议采用专属政策的数据类型（改编自 [50]）

11. 列出要求采用专属政策的数据类型（改编自 [50]）

12. 代码共享要求

（a）未提及代码共享 （b）鼓励代码共享，但非强制要求 （c）要求代码共享

13. 代码可复现性

(a) 未提及依赖项 (b) 鼓励列出依赖项和版本 (c) 要求列出依赖项和版本 (d) 要求提供容器/安装脚本 (e) 要求通过容器/脚本实现图表的完全可复现性

14. 版本管理和持久标识符

(a) 未提及持久标识符或版本 (b) 鼓励提供持久标识符或版本 (c) 要求提供持久标识符或版本

15. 代码质量标准

（a）未提及代码质量标准 （b）鼓励遵循标准 （c）要求遵循标准 （d）针对代码质量为审稿人制定具体评审标准

16. 自动化测试

(a) 未提及自动化测试 (b) 鼓励进行自动化测试 (c) 必须进行自动化测试

17. 代码文档

(a) 未提及文档编制标准 (b) 鼓励编制文档 (c) 必须编制文档

18. 代码检查标准

(a) 未提及代码检查标准 (b) 鼓励遵守代码检查标准 (c) 必须遵守代码检查标准

19. 代码开发与持续集成

(a) 未提及开发标准 (b) 鼓励采用开发标准 (c) 要求采用开发标准

VI. 致谢

本文/发表成果是 COST 行动 CA22154——面向功能材料工程的数据驱动应用：开放网络（DAEMON）——中的联合工作成果，该行动由 COST（欧洲科学技术合作组织）支持。JS 获得了欧洲研究委员会（ERC）的支持，资助来自欧盟地平线2020研究与创新计划项目 HERO（资助协议编号：810451），以及编号为 233444 的瑞士国家科学基金会 Ambizione 资助。LH 获得了英国研究与创新署地平线资助（玛丽·斯克沃多夫斯卡-居里行动，EP/Y024923/1）的支持。KB 感谢塞尔维亚共和国科学、技术与创新部通过合同编号 451-03-33/2026-03/200017 提供的支持。MS 感谢欧盟通过欧洲研究委员会资助项目（项目编号：101161287）提供的支持。然而，本文所表达的观点和意见仅代表作者本人，并不一定反映欧盟或欧洲研究委员会执行机构的观点和意见。欧盟和资助机构均不对这些观点和意见承担责任。SB 感谢波兰国家科学中心通过 SONATA BIS 资助 DEC-2023/50/E/ST3/00569，以及波兰科学基金会

通过 FIRST TEAM FENG.02.02-IP.05-0177/23 项目提供的支持。所有作者衷心感谢 Nicola Spaldin、Julian Dederke 和 Nicolai Bissantz 富有帮助且富有洞见的讨论，并感谢 ETHLibrary 提供的支持。

VII. 作者贡献

概念化：FG、HM、JS、LH、MS、RKSAH、SG；方法论：HM、JS、LH、MS、SG；软件：HM、IL、JS、LH、MS、NL、SG；验证：FG、JS、LH、MS、RKSAH、SG；形式分析：FG、HM、JS、LH、MS、SG；调查：BK、ES、FG、HM、IL、JMO、JS、KB、KK、KR、LH、MS、MT、PP、PR、RKSAH、SB、SG、SPS；数据整理：FG、HM、JS、LH、NL、SG；撰写——初稿：FG、HM、JS、LH、MS、SG；撰写——审稿与编辑：BK、ES、FG、HM、IL、JMO、JS、KB、KK、KR、LH、MS、MT、PP、PR、RKSAH、SB、SG、SPS；可视化：FG、HM、JS、LH、MS；监督：FG、JS、LH、MS；项目管理：JS、LH、MS；资金筹措：FG、KR、MT

VIII. 利益冲突

KR 是《Journal of Physics:Materials》（IOP Publishing）编辑委员会成员

IX. 数据和代码可用性

所有用于复现该研究的数据和代码均可在 https://github.com/daemoncost/Journal-Research-Data-Policy 获取。发表时的代码版本将镜像存档至 Zenodo，并分配持久标识符。

[1]I. Tanaka、K. Rajan 和 C. Wolverton，面向材料创新的数据中心科学，《MRS Bulletin》**43**，659（2018）。[2] S. Sandfeld、T. Dahmen、F. O. R. Fischer、C. Eberl、S. Klein、M. Selzer、B. Nestler、J. Möller、F. Mücklich、M. Engstler、S. Diebels、R. Tschuncky、A. Prakash、D. Steinberger、C. Kübel、H.-G. Herrmann 和 R. Schubotz，《材料科学与工程中的数字化转型》，德国材料学会（2018）。[3] M. Scheffler、M. Aeschlimann、M. Albrecht、T. Bereau、H.-J. Bungartz、C. Felser、M. Greiner、A. Groß、C. T. Koch、K. Kremer，等，《FAIR数据为材料研究开辟新视野》，《自然》**604**，635（2022）。[4] S. Bauer、P. Benner、T. Bereau、V. Blum、M. Boley、C. Carbogno、R. Catlow、G. Dehm、S. Eibl、R. Ernstorfer、Á. Fekete、L. Foppa、P. Fratzl、C. Freysoldt、B. Gault、L. M. Ghiringhelli、S. K. Giri、A. Gladyshev、P. Goyal、J. Hattrick-Simpers、L. Kabalan、P. Karpov、M. S. Khorrami、C. T. Koch、S. Kokott、T. Kosch、I. Kowalec、K. Kremer、A. Leitherer、Y. Li、C. H. Liebscher、A. Logsdail、Z. Lu、F. Luong、A. Marek、F. Merz、J. Rezaei Mianroodi、J. Neugebauer、Z. Pei、T. A. R. Purcell、D. Raabe、M. Rampp、M. Rossi、J. M. M. Rost、J. E. Saal、U. Saalmann、K. N. Sasidhar、A. Saxena、L. Sbailo、M. Scheidgen、M. Schloz、D. Schmidt、S. Teshuva、A. Trunschke、Y. Wei、G. Weikum、R. P. Xian、Y. Yao、J. Yin、M. Zhao 和 M. Scheffler，《数据中心材料科学路线图》，Model. Simul. Mater. Sci. Eng. **32**，063301（2024）。[5] O. Kononova、T. He、H. Huo、A. Trewartha、E. A. Olivetti 和 G. Ceder，《材料研究中文本挖掘的机遇与挑战》，iScience **24**，102155（2021）。[6] M. Schilling-Wilhelmi、M. Ríos-García、S. Shabih、M. V. Gil、S. Miret、C. T. Koch、J. A. Márquez 和 K. M. Jablonka，《从文本到洞见：用于化学数据提取的大语言模型》，Chem. Soc. Rev. **54**，1125（2025）。[7] M. D. Wilkinson、M. Dumontier、I. J. Aalbersberg、G. Appleton、M. Axton、A. Baak、N. Blomberg、J.-W. Boiten、L. B. da Silva Santos、P. E. Bourne、J. Bouw- man、A. J. Brookes、T. Clark、M. Crosas、I. Dillo、O. Du- mon、S. Edmunds、C. T. Evelo、R. Finkers、A. Gonzalez- Beltran、A. J. G. Gray、P. Groth、C. Goble、J. S. Grethe、J. Heringa、P. A. C. 't Hoen、R. Hooft、T. Kuhn、R. Kok、J. Kok、S. J. Lusher、M. E. Martone、A. Mons、A. L. Packer、B. Persson、P. Rocca-Serra、M. Roos、R. van Schaik、S.-A. Sansone、E. Schultes、T. Sengstag、T. Slater、G. Strawn、M. A. Swertz、M. Thompson、J. van der Lei、E. van Mulligen、J. Velterop、A. Waag- meester、P. Wittenburg、K. Wolstencroft、J. Zhao 和 B. Mons，《科学数据管理与管理维护的 FAIR 指导原则》，《科学数据》**3**，160018（2016年）。[8] D. Donoho，《奇点处的数据科学》（2023年），arXiv:2310.00865。[9]R. Ramachandran、K. Bugbee 和 K. Murphy，《从开放数据到开放科学》，《地球与空间科学》**8**，e2020EA001562（2021年）。[10] I. Shumailov、Z. Shumaylov、Y. Zhao、N. Papernot、R. Anderson 和 Y. Gal，《人工智能模型在使用递归生成的数据训练时会发生崩溃》，《自然》**631**，755（2024年）。[11] 晶体学：蛋白质数据库，《自然新生物学》**233**，223（1971年）。[12] F. C. Bernstein、T. F. Koetzle、G. J. Williams、E. F. M. Jr.、M. D. Brice、J. R. Rodgers、O. Kennard、T. Shi- manouchi 和 M. Tasumi，《蛋白质数据库：大分子结构的基于计算机的档案文件》，Journal of Molecular Biology **112**，535（1977年）。[13] J. D. Westbrook、J. Y. Young、C. Shao、Z. Feng、V. Guranovic、C. L. Lawson、B. Vallat、P. D. Adams、J. M. Berrisford、G. Bricogne、K. Diederichs、R. P. Joosten、P. Keller、N. W. Moriarty、O. V. Sobolev、S. Velankar、C. Vonrhein、D. G. Waterman、G. Kurisu、H. M. Berman、S. K. Burley 和 E. Peisach，PDBx/mmCIF生态系统：

结构生物学的基础语义工具，Journal of Molecular Biology **434**，167599（2022年）。[14] T. Terwilliger 和 G. Bricogne，《大分子结构模型与X射线晶体学软件的持续相互改进：已提交实验数据的双重作用》，Acta Crystallographica D **70**，2533 2014( )。[15] wwPDB，*wwPDB*政策与处理流程文件。*A*节：*wwPDB*政策，技术报告（wwPDB，2024年）。[16] M. Varadi、D. Bertoni、P. Magana、U. Paramval、I. Pidruchna、M. Radhakrishnan、M. Tsenkov、S. Nair、M. Mirdita、J. Yeo、O. Kovalevskiy、K. Tunyasuvu- nakool、A. Laydon、A. Žídek、H. Tomlinson、D. Hari- haran、J. Abrahamson、T. Green、J. Jumper、E. Birney、M. Steinegger、D. Hassabis 和 S. Velankar，《2024 年 AlphaFold 蛋白质结构数据库：为超过 2.14亿条蛋白质序列提供结构覆盖》，《核酸研究》**52**，D368（2024年）。[17] J. Jumper、R. Evans、A. Pritzel、T. Green、M. Fig- urnov、O. Ronneberger、K. Tunyasuvunakool、R. Bates、A. Žídek、A. Potapenko、等，《借助 AlphaFold 高精度预测蛋白质结构》，《自然》**596**，583（2021年）。[18] M. Varadi、D. Bertoni、P. Magana、U. Paramval、I. Pidruchna、M. Radhakrishnan、M. Tsenkov、S. Nair、M. Mirdita、J. Yeo、O. Kovalevskiy、K. Tunyasuvu- nakool、A. Laydon、A. Žídek、H. Tomlinson、D. Hari- haran、J. Abrahamson、T. Green、J. Jumper、E. Birney、M. Steinegger、D. Hassabis 和 S. Velankar，《2024年 AlphaFold 蛋白质结构数据库：为超过2.14亿条蛋白质序列提供结构覆盖》，《核酸研究》**52**，D368–D375（2023年）。[19] F. Zhao 和 J. Xu，《用于蛋白质结构与功能研究的位置特异性距离依赖统计势》，Structure **20**，1118（2012年）。[20] N. Methods，《数据共享进入结构生物学领域》，《自然·方法》**13**，381（2016年）。[21] M. Barker、N. P. Chue Hong、D. S. Katz、A.-L. Lam- precht、C. Martinez-Ortiz、F. Psomopoulos、J. Harrow、L. J. Castro、M. Gruenpeter、P. A. Martinez 和 T. Hon- eyman，《为研究软件提出 FAIR 原则》，《科学数据》**9**，622（2022年）。[22] L. M. Ghiringhelli、C. Baldauf、T. Bereau、S. Brock- hauser、C. Carbogno、J. Chamanara、S. Cozzini、S. Cur- tarolo、C. Draxl、S. Dwaraknath、Á. Fekete、J. Kermode、C. T. Koch、M. Kühbach、A. N. Ladines、P. Lambrix、M.- O. Himmer、S. V. Levchenko、M. Oliveira、A. Michalchuk、R. E. Miller、B. Onat、P. Pavone、G. Pizzi、B. Re- gler、G.-M. Rignanese、J. Schaarschmidt、M. Scheid- gen、A. Schneidewind、T. Sheveleva、C. Su、D. Usvyat、O. Valsson、C. Wöll 和 M. Scheffler，《面向数据驱动材料科学的共享元数据》，《科学数据》**10**，626（2023年）。[23] S. Curtarolo、W. Setyawan、G. L. Hart、M. Jahnatek、R. V. Chepulskii、R. H. Taylor、S. Wang、J. Xue、K. Yang、O. Levy、M. J. Mehl、H. T. Stokes、D. O. Demchenko 和 D. Morgan，《AFLOW：用于高通量材料发现的自动化框架》，《计算材料科学》**58**，218 2012( )。[24] A. Jain、S. P. Ong、G. Hautier、W. Chen、W. D. Richards、S. Dacek、S. Cholia、D. Gunter、D. Skinner、G. Ceder 和 K. A. Persson，评论：《材料项目：一种加速材料创新，APL Mater.**1**，011002（2013）。[25] S. Kirklin、J. E. Saal、B. Meredig、A. Thompson、J. W. Doak、M. Aykol、S. Rühl 和 C. Wolverton，《开放量子材料数据库》（OQMD）：评估DFT形成能的准确性，npj Comput. Mater. **1**，15010（2015年）。[26] C. Draxl 和 M. Scheffler，NOMAD实验室：从数据共享到人工智能，J. Phys.: Mater. **2**，036001（2019）。[27] T. Cavignac、J. Schmidt、P.-P. De Breuck、A. Loew、T. F. T. Cerqueira、H.-C. Wang、A. Bochkarev、Y. Lyso- gorskiy、A. H. Romero、R. Drautz、S. Botti 和 M. A. L. Marques，由人工智能驱动的Alexandria数据库扩展与应用（2025年）。[28] L. Talirz、S. Kumbhar、E. Passaro、A. V. Yakutovich、V. Granata、F. Gargiulo、M. Borelli、M. Uhrin、S. P. Huber、S. Zoupanos、C. S. Adorf、C. W. Andersen、O. Schütt、C. A. Pignedoli、D. Passerone、J. VandeVon- dele、T. C. Schulthess、B. Smit、G. Pizzi 和 N. Marzari，Materials Clo ud：开放计算科学平台，《科学数据》**7**，299（2020）。[29] M. L. Evans、J. Bergsma、A. Merkys、C. W. Andersen、O. B. Andersson、D. Beltrán、E. Blokhin、T. M. Boland、R. Castañeda Balderas、K. Choudhary、A. Díaz Díaz、R. Domínguez García、H. Eckert、K. Eimre、M. E. Fuentes Montero、A. M. Krajewski、J. J. Mortensen、J. M. Nápoles Duarte、J. Pietryga、J. Qi、F. d. J. Trejo Carrillo、A. Vaitkus、J. Yu、A. Zettel、P. B. de Castro、J. Carlsson、T. F. T. Cerqueira、S. Divilov、H. Hajiyani、F. Hanke、K. Jose、C. Oses、J. Riebesell、J. Schmidt、D. Winston、C. Xie、X. Yang、S. Bonella、S. Botti、S. Curtarolo、C. Draxl、L. E. Fuentes Cobas、A. Hospital、Z.-K. Liu、M. A. L. Marques、N. Marzari、A. J. Morris、S. P. Ong、M. Orozco、K. A. Persson、K. S. Thygesen、C. Wolverton、M. Scheidgen、C. Toher、G. J. Conduit、G. Pizzi、S. Gražulis、G.-M. Rignanese 和 R. Armiento，用于材料发现、设计和数据交换的OPTIMADE应用程序接口：进展与应用，Digital Discovery **3**，1509（2024年）。[30] L. Ward、S. Babinec、E. J. Dufek、D. A. Howey、V. Viswanathan、M. Aykol、D. A. Beck、B. Blaiszik、B.-R. Chen、G. Crabtree、S. Clark、V. De Angelis、P. Dechent、M. Dubarry、E. E. Eggleton、D. P. Finegan、I. Foster、C. B. Gopal、P. K. Herring、V. W. Hu、N. H. Paulson、Y. Preger、D. Uwe-Sauer、K. Smith、S. W. Snyder、S. Sri- pad、T. R. Tanim 和 L. Teo，电池数据基因组的原则，Joule **6**，2253（2022）。[31] A. White，材料基因组计划：一年之后，MRS Bulletin **37**，715（2012）。[32]材料研究数据联盟（MaRDA），主页，https://www.marda-alliance.org/（2025），访问日期：2025-08-15。[33]欧洲研究委员会，开放研究数据与数据管理计划，https://erc.europa. eu/sites/default/files/document/file/ERC_ info_document-Open_Research_Data_and_Data_ Management_ Plans.pdf（2022年）。[34]德国研究联合会（DFG），研究数据处理，https://www.dfg.de/ en/research-funding/funding-initiative/ research-data（2025）。[35] 瑞士国家科学基金会，开放研究数据，https://www.snf.ch/en/dMILj9t4LNk8NwyR/topic/ open-research-data（2025）。

[36]克罗地亚科学基金会，研究数据管理计划，https://hrzz.hr/plan-upravljanja-istrazivackim-podacima-za-projek

t克罗地亚科学基金会的

(2022)。[37]冰岛研究中心，《科学出版物：开放获取政策》，https://en.rannis.is/ activities/open-access/ (2025 18, 2025)，访问日期：6月。[38]主题数字能力中心（TDCC），自然与工程科学（tdcc-nes），https://tdcc.nl/ about-tddc/nes/ (2025)，访问日期：2026-02-02。[39]波鸿鲁尔大学，《研究数据管理指南》，https://researchdata.ruhr-uni-bochum.de/wp-content/uploads/2024/07/LeitlinienFDM.pdf (2018)。[40] 维尔纽斯大学，《维尔纽斯大学开放科学政策指南》，https://www.vu.lt/site_files/Vertimai/ EN_Translation_VU_atvirojo_mokslo_politikos_ gaires.pdf (2022)。[41] A. E. Mansour、L. Rotheray、K. Helbig、S. Botti、H. B. Weber、M. Aeschlimann 和 C. Draxl，《Fairmat 指南：如何撰写数据管理计划——面向凝聚态物理和材料科学界的实用指南》，《研究数据基础设施会议论文集》**1**，10.52825/cordi.v1i.378 (2023)。[42] L. Himanen、A. Geurts、A. S. Foster 和 P. Rinke，《数据驱动的材料科学：现状、挑战与展望》，Adv. Sci. **6**，1900808 (2019)。[43] D. G. E. Gomes、P. Pottier、R. Crystal-Ornelas、E. J. Hudgins、V. Foroughirad、L. L. Sánchez-Reyes、R. Turba、P. A. Martinez、D. Moreau、M. G. Bertram、C. A. Smout 和 K. M. Gaynor，《我们为什么不共享数据和代码？公众存档实践的感知障碍与益处》，Proc. Biol. Sci. **289**，20221113 (2022)。[44] N. Hartl、E. Wössner 和 Y. Sure-Vetter，《国家研究数据基础设施（NFDI）》，Informatik Spektrum **44**，370 (2021)。[45] Y. Kim 和 J. M. Stanton，《影响科学家数据共享行为的机构和个人因素：一项多层次分析》，J. Assn. Inf. Sci. Tec. **67**，776 (2016)。[46] V. Stodden、P. Guo 和 Z. Ma，《迈向可复现计算研究：对期刊数据与代码政策采纳情况的实证分析》，PLOS ONE **8**，e67111 (2013)。[47] N. A. Vasilevsky、J. Minnier、M. A. Haendel 和 R. E. Champieux，《可复现与可重复利用的研究：期刊数据共享政策是否达标？》，PeerJ **5**，e3208 (2017)。[48] K. W. McCain，《强制共享：自然科学领域的期刊政策》，《科学传播》**16**，403 (1995)。[49] C. Barbui，《共享各种类型的临床数据并协调期刊标准》，BMC Med. **14**，63 (2016)。[50] D. B. Resnik、M. Morales、R. Landrum、M. Shi、J. Minnier、N. A. Vasilevsky 和 R. E. Champieux，《影响因子和学科对期刊数据共享政策的影响》，Accountability in Research **26**，139 (2019)。[51] A. M. Rousi 和 M. Laakso，《期刊研究数据共享政策：对神经科学、物理学和运筹学领域高被引期刊的研究》，《科学计量学》**124**，131 (2020)。[52] M. Crosas、J. Gautier、S. Karcher、D. Kirilova、G. Otalora 和 A. Schwartz，《社会科学高排名期刊的数据政策》，SocArXiv 预印本 (2018)。

[53] E. Castro、M. Crosas、A. Garnett、K. Sheridan 和 M. Altman，《评估并推动开放获取期刊中的开放数据实践》，学术出版期刊**49**，66（2017年）。[54] L. Naughton 和 D. Kernohan，《英国期刊》的见解，UKSG**29**，84（2016年）。[55] B. Blahous、J. Gorraiz、C. Gumpenberger、O. Lehner 和 U. Ulrych，《审视期刊的数据政策：其力度、范围和影响》，Bibliometrie-Praxis und Forschung**5**，10.5283/BPF.269（2016年）。[56] J. Herndon 和 R. O'Reilly，《社会科学学术期刊的数据共享政策：作为学术交流形式的数据共享期望不断演变》，载于《数据馆员学：理论与实践中的学术数据馆员》，L. Kellam 和 K. Thompson 编（美国图书馆协会，芝加哥，2016年），第219–242页。[57] P. Sturges、M. Bamkin、J. H. Anders、B. Hubbard、A. Hussain 和 M. Heeley，《研究数据共享：制定利益相关者驱动的期刊政策模型》，信息科学与技术协会期刊**66**，2445（2015年）。[58] W. Zenk-Möltgen 和 G. Lepthien，《社会学期刊中的数据共享》，Online Information Review**38**，709（2014年）。[59] N. Moles，《Data-pe：评估学术期刊数据出版政策的框架》，Data Science Journal**13**，192（2015年）。[60] s. gherghina 和 a. katsanidou，《政治学期刊中的数据可用性》，European Political Science**12**，333（2013年）。[61] N. M. Weber、H. A. Piwowar 和 T. J. Vision，《评估环境科学中的数据引用与共享政策：评估环境科学中的数据引用与共享政策》，美国信息科学与技术学会会议录**47**，1（2010年）。[62] H. Piwowar 和 W. Chapman，《研究数据共享期刊政策综述》，Nature Precedings，10.1038/npre.2008.1700.1（2008年）。[63] R. Grant 和 I. Hrynaszkiewicz，《在《自然》期刊中引入数据可用性声明对作者和编辑的影响》，载于国际数字策展会议论文集（2018年），已提交至《国际数字策展期刊》。[64] B. Suhr、J. Dungl 和 A. Stocker，《材料科学与工程领域研究数据的检索、再利用与共享——一项定性访谈研究》，PLOS ONE**15**，1（2020年）。[65] M. Houillon、J. Klar、Z. Boutanios、T. Stary、T. Cojean、H. Anzt 和 A. Loewe，《FACILE-RS：轻松归档和长期保存研究软件存储库》，Journal of Open Source Software**10**，7330（2025年）。[66] D. Science，Dimensions API，https://api-lab.dimensions.ai/，访问日期：2024年4月。[67] M. Gabelica、R. Bojčić 和 L. Puljak，《许多研究人员未遵守其已发布的数据共享声明：一项混合方法研究》，J. Clin. Epidemiol.**150**，33（2022年）。[68] D. G. Hamilton、M. J. Page、S. Finch、S. Everitt 和 F. Fidler，《癌症研究人员多常公开其数据和代码？哪些因素与共享相关？》，BMC Med.**20**，438（2022年）。

[69] Parasoft，汽车行业中的 *ISO 26262* 软件合规性，技术报告（Parasoft，2025年）。[70] 美国国家航空航天局，*NASA* 软件安全指南，标准 NASA-GB-8719.13（美国国家航空航天局，2004年）。[71] L. N. Joppa、G. McInerny、R. Harper、L. Salido、K. Takeda、K. O'Hara、D. Gavaghan 和 S. Emmott，《科学软件使用中的令人担忧趋势》，《科学》**340**，814（2013年）。[72] J. Brase，《DataCite——全球研究数据注册机构》，载于*2009*年第四届科学技术信息资源合作与推广国际会议（2009年），第257–261页。[73] J. Neumann 和 J. Brase，《DataCite 与研究数据 DOI 名称》，J. Comput.-Aided Mol. Des.**28**，1035（2014年）。[74] M. Barker、N. P. Chue Hong、D. S. Katz、A.-L. Lamprecht、C. Martinez-Ortiz、F. Psomopoulos、J. Harrow、L. J. Castro、M. Gruenpeter、P. A. Martinez 和 T. Honeyman，《介绍研究软件 FAIR 原则》，《科学数据》**9**，622（2022年）。[75] T. Fischer、W. Vollprecht、B. Zalmstra、R. Arts、T. de Jager、A. Fontan、A. D. Hines、M. Milford、S. Traversaro、D. Claes 和 S. Raine，《Pixi：面向机器人学和人工智能的统一软件开发与分发》（2025年）。[76] 欧洲核子研究组织和 OpenAIRE，Zenodo（2013年）。[77] G. Wilson、D. A. Aruliah、C. T. Brown、N. P. Chue Hong、M. Davis、R. T. Guy、S. H. D. Haddock、K. D. Huff、I. M. Mitchell、M. D. Plumbley、B. Waugh、E. P. White 和 P. Wilson，《科学计算最佳实践》，《公共科学图书馆·生物学》**12**，e1001745（2014年）。[78] R. C. Jiménez, M. Kuzak, M. Alhamdoosh, M. Barker, B. Batut, M. Borg, S. Capella-Gutierrez, N. C. Hong, M. Cook, M. Corpas, M. Flannery, L. Garcia, J. Ll. Gelpí, S. Gladman, C. Goble, M. G. Ferreiro, A. Gonzalez- Beltran, P. C. Griffin, B. Grüning, J. Hagberg, P. Holub, R. Hooft, J. Ison, D. S. Katz, B. Leskošek, F. L. Gómez, L. J. Oliveira, D. Mellor, R. Mosbergen, N. Mul- der, Y. Perez-Riverol, R. Pergl, H. Pichler, B. Pope, F. Sanz, M. V. Schneider, V. Stodden, R. Suchecki, R. S. Vařeková, H.-A. Talvik, I. Todorov, A. Treloar, S. Tyagi, M. van Gompel, D. Vaughan, A. Via, X. Wang, N. S. Watson-Haigh, and S. Crouch, 《促进研究软件最佳实践的四项简单建议》，F1000Research **6**, 876 (2017). [79] N. Blumenröhr, P.-J. Ost, F. Kraus, and A. Streit, 《通过 FAIR 数字对象实现全球协调的数据空间》，载于*2024*年 *IEEE* 国际大数据会议 *(BigData)* (IEEE, 2024年) pp. 374–383. [80] S. Soiland-Reyes, C. Goble, and P. Groth, 《评估 FAIR 数字对象和关联数据作为分布式对象系统》，《PeerJ 计算机科学》**10**, e1781 (2024年). [81] I. Anders, C. Blanchi, D. Broder, M. Hellström, S. I s- lam, T. Jejkal, L. Lannom, K. P. von Gehlen, R. Quick, A. Schlemmer, U. Schwardmann, S. Soiland-Reyes, S. George, D. van Uytvanck, C. Weiland, P. Wittenburg, and C. Zwölf, 《FDO 论坛 FDO 要求规范》，v3.0, Zenodo 10.5281/ZENODO.7782262 (2023年). [82] H. Koers, D. Bangert, E. Hermans, R. van Horik, M. de Jong, and M. Mokrane, 《FAIR 数据生态系统服务建议》，《Patterns》**1**, 100058 (2020年). [83] The Carpentries (2025年), [在线；访问日期：11月20日 2025]. [84] CodeRefinery (2025年), [在线；访问日期：11月20日 2025].