



# 公开研究数据，保障学术研究的可检验性和可重复性

《现代图书情报技术》编辑部\*

(中国科学院文献情报中心 北京 100190)

学术研究成果必须通过可靠的证据来证明其科学性、可靠性，来支持研究结果的重复验证。但是，许多研究依赖大量、复杂的数据收集、处理、分析、表达，同行评议专家和读者需要了解支撑研究结论的关键数据及其收集、处理、分析和表达过程，否则无法对研究结论的科学性和可靠性做出判断，无法可靠进行重复验证。所以，英国皇家学会明确提出<sup>[1]</sup>，要保证科学知识本身的可检验、可重复和可利用，就必须从源头起打破数据鸿沟，保证科学数据的可获得、可理解、可评估和可利用。因此，论文投稿时提交和发表时公布支撑数据，已成为越来越多学术期刊的基本要求，例如 Science、Nature、Cell、PLOS Biology 等。

为了保证学术水平、保障客观检验、支持重复验证、提高科学诚信，《现代图书情报技术》将从 2016 年第 2 期起，将要求所有发表论文提交支撑论文结论的数据，并通过适当方式供研究共同体或社会公众共享。

## 1 提交什么数据？

1.1 根据美国国家信息标准组织《关于期刊论文支撑资料的推荐实践》<sup>[2]</sup>，支撑论文结论的数据可分为两类：

(1) 内在数据(Integral Content)，即那些对于本领域读者了解和验证论文研究过程和研究内容所必不可少的，但因为技术、内容或版面组织、呈现平台组织等原因没有纳入论文正文的数据，例如研究方法和实施过程介绍，原始研究量表，详细结果数据，理解和验证结论所需要的与研究结果直接相关的图表或多媒体数据；

(2) 附加数据(Additional Content)，即那些有助于本领域读者及其他领域读者进一步了解研究环境、研究过程和研究结果细节或操作的相关的或扩展的文字、图表、数据、多媒体等数据，例如关于研究方法、工具或设施设备细节、扩展的研究数据、扩展的参考文献等。

1.2 根据本刊报道范围和论文类型，以下类型为内在数据，为同行评议专家和本领域研究人员检验研究结论所必须，作者必须提供：

(1) 研究方法详细描述，包括抽样方法、实验环境配置、应用环境配置、实验操作过程等，其中要点应在正文中提供，详细信息作为数据提供；

(2) 直接研究工具，例如问卷、访谈提纲、测试题目、计算算法、样本数据等，其中关键数据应在正文中提供和分析，但完整的数据必须作为数据提供；

(3) 直接研究结果数据，包括直接产生正文图表和正文所描述结论的样本数据表、测试数据记录、调查或观察数据记录、计算或模拟数据记录等。

1.3 根据本刊报道范围和论文类型，以下类型为附加数据，有助于本领域科研人员和其他读者进一步理解研究结论的，作者可以提供：

(1) 研究方法背景信息，例如关于所使用的主要方法、技术、工具等的进一步详细描述，其中第三方的内容可以通过具体直接的参考文献来提供；

\* 执笔：张晓林、彭希珺。

(2) 扩展的研究结果数据, 例如关于研究结论的进一步的原始实验数据, 它们可能没有直接用于产生正文图表等、但与理解和支持正文研究结论有关;

(3) 扩展的参考文献列表, 例如对于帮助读者理解研究方法、研究过程和研究结论有帮助的参考文献。

## 2 如何提交?

2.1 作者在提交论文初稿时必须通过期刊投稿系统提交支撑数据, 或提交关于支撑数据获取方式的明细说明。作者应在正文结尾以数据清单形式明确列出所提交的每个数据集, 给出数据集逻辑名称、文件名称、文件类型、建议开放时滞期、建议使用许可等。

2.2 参照 NIH<sup>[3]</sup>、Nature<sup>[4]</sup>和 PLOS<sup>[5]</sup>的数据提交政策, 本刊允许作者以下列形式提交数据(按优先顺序排列):

(1) 将数据文件存储在可公共获取的数据存储库中, 例如 SciDB 数据库<sup>[6]</sup>。作者必须向编辑部提供这些数据文件在相应系统中的唯一标识号、URL 以及相应的获取流程说明;

(2) 以附属数据文件形式向编辑部提供;

(3) 如果数据文件已经存储在可供第三方或公众获取的其他系统, 作者可提供数据文件在这些系统中的标识号、URL 以及相应的获取流程说明;

(4) 作者承诺可向索取数据的本学科研究者、具有合理研究或学习理由的其他学科研究者及公众提供数据文件。作者必须向编辑部提供关于数据文件的索取条件和流程说明。

鼓励作者以同一方式提供论文的所有支撑数据。

2.3 作者在提供数据时, 应承诺作者有合法权利提交相关数据, 应承诺数据没有违反著作权法、国家保密规定、资助单位或所在单位的合法管理规定。

2.4 无论是哪种形式提供, 数据应符合以下技术要求:

(1) 以通用的计算机可读数据格式提供, 例如规范的常用的数据格式、数据库格式、Excel 数据表等; 如果数据以 XML 文档格式提供, 必须给出 XML 格式文件或其可靠 URL; 一般不能以文本格式或 PDF 格式提供。如果采用非常规格式进行数据处理, 作者应该将原始数据不失真地转换为常规格式。如果提供非常规格式的数据, 作者应提供相应的数据读取软件或其公共获取途径;

(2) 提供关于步骤(1)中所提供数据集的所有数据元素的元数据描述。这些描述可以以文本格式或 PDF 格式提供, 如果元数据已包括在 XML 格式的数据文件中, 作者应予明确说明;

(3) 提供多媒体内容时, 必须以规范的常见的数据格式提供, 并提供相应的关于数据格式的说明文字文档。

2.5 如果作者将数据存储于可公共获取的数据存储库中, 这些数据存储库必须符合以下条件: 得到该领域科研群体认可, 能长期稳定运行, 提供专业化数据监护, 提供规范的数据使用权益许可, 支持国际或学科领域通用的数据集唯一标识符, 支持同行评议专家对数据集的匿名评审, 免费或仅以不超过存储管理成本收取存储费用, 免费或仅以不超过存储管理成本提供获取服务。

## 3 如何共享?

(1) 作者必须允许同行评议专家和编辑部在收到稿件后即可获取数据。如果数据存储于公共数据存储库或其他系统且尚未向公众开放时, 作者必须保证评议专家和编辑部能合法、便利地获取这些数据。

(2) 论文如果被录用, 在发表时应参考文献引用著录方式, 在论文正文内容中首次提及数据内容的位置“引用”支撑数据, 数据引用方式采用 DataCite 元数据格式<sup>[7]</sup>, 最低限度必须采用所有必备元素, 一般应采用所有必备元素和尽可能多的推荐元素; 作者提供的所有支撑数据文件清单在论文正文后和参考文献前以“支撑数据”名义列出; 网络版正文中将同时给出各个支撑数据集的链接。

(3) 论文发表后, 如果支撑数据文件清单中的数据集没有在期刊网站、公共获取数据存储库中或其他系统中发布, 或者仍在这些系统的开放时滞期内, 作者应允许本学科研究者立即向作者索取相应的数据集。

(4) 作者应允许在论文发表后不超过 6 个月时间内, 本学科研究者通过期刊网站、公共获取的数据存储库或其他系统可开放获取支撑数据; 作者应允许在论文发表后不超过一年时间内, 公众通过期刊网站、公共获取的数

据存储库或其他系统可开放获取支撑数据。编辑部鼓励论文发表时即提供数据开放获取。

(5) 作者可根据开放获取的原则对开放获取的支撑数据提出使用许可,原则上应采用 Open Data Common 许可协议<sup>[8]</sup>。

(6) 作者向除同行评议专家和编辑部外的其他人员提供数据时,应采取通行的规范措施来保护受试对象的隐私信息。因此,作者在投稿时,可以在不影响同行评议专家和编辑部评审论文的情况下对数据进行匿名化处理;如果匿名化处理可能影响评审,作者须要同时提供没有匿名化和经过匿名化处理的两个版本。同行评议专家和编辑部承诺保护受试对象的隐私信息。

(7) 所有共享支撑数据的人员和系统承诺按照法律和许可使用数据。

#### 4 编辑部对数据如何管理?

(1) 编辑部在接受投稿时审查是否提供了内在数据、所提供数据的描述是否清楚、所提供的数据文件是否可获取。如果其中一个答案为否,编辑部将要求作者补充提供,在获得符合条件的数据后才进入同行评议环节。

(2) 同行评议专家在评审论文时将根据论文研究方法和内容评审的需要检查所需要提供的支撑数据是否提供、是否描述清晰、是否可获取。同行评议专家可能要求作者补充提供数据,作者应及时提供。

(3) 作者对所提供支撑数据的真实性、完整性、可靠性负责。同行评议专家和编辑部主要对支撑数据与论文内容的相关性、数据文档描述的完备性、数据文档的可获得性等进行检查。

(4) 同行评议专家和编辑部承诺不向任何第三方提供用于评审的数据,承诺保护作者及其他数据权利人的合法利益。

(5) 如果作者以附属文件形式提供支撑数据,编辑部将通过期刊网站专门的数据管理模块保存数据和在网络版正文中提供数据链接。编辑部承诺对数据进行专业化监护。如果作者按照前述“3(4)”设置开放获取时滞期,编辑部将尊重作者要求实施时滞期管理。

(6) 编辑部在征得作者同意后,可将作者提供的数据向符合条件的公共获取的数据存储库转移,以支持更为高效方便的数据服务。

编辑部将持续评估自己的研究数据公开政策及其实施流程。

#### 参考文献:

- [1] Royal Society. Science as an Open Enterprise [EB/OL]. [2015-09-24]. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/>.
- [2] NISO RP-15-2013 Recommended Practices for Online Supplemental Journal Article Materials [EB/OL]. [2015-09-24]. <http://www.niso.org/publications/rp/rp-15-2013>.
- [3] NIH Data Sharing Policies [EB/OL]. [2015-09-24]. [http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/).
- [4] NPG Availability of Data, Materials and Methods [EB/OL]. [2015-09-24]. <http://www.nature.com/authors/policies/availability.html>.
- [5] PLOS Data Availability [EB/OL]. [2015-09-24]. <http://journals.plos.org/plosone/s/data-availability>.
- [6] SciDB 科学数据存储库[EB/OL]. [2015-09-24]. <http://www.scidb.cn/index>. (Science Data Bank [EB/OL]. [2015-09-24]. <http://www.scidb.cn/index>.)
- [7] DataCite Metadata Schema for the Publication and Citation of Research Data. V3.1 [EB/OL]. [2015-09-24]. [http://schema.datacite.org/meta/kernel-3.1/doc/DataCite-MetadataKernel\\_v3.1.pdf](http://schema.datacite.org/meta/kernel-3.1/doc/DataCite-MetadataKernel_v3.1.pdf).
- [8] Open Data Commons Licenses [EB/OL]. [2015-09-24]. <http://opendatacommons.org/licenses/>.