

• 医学信息组织与利用 •

WHO 西太平洋地区医学索引收录期刊的数据提交能力分析

方 安 杨晨柳 王军辉

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 简要介绍 WHO 西太平洋地区医学索引 (WPRIM) 和当前 WPRIM 数据采集的途径及存在的问题, 基于问卷调查, 对 WPRIM 中国期刊从期刊信息数据提交能力、提交各方式倾向性和工作建议等方面进行分析, 针对调查结果所反映的问题进行讨论, 包括改进调查方式, 提高 WPRIM 数据完整性, 提升编辑部数据提交能力, 推进 WPRIM 宣传工作等。

〔关键词〕 西太平洋地区医学索引; 全球卫生图书馆; 全球医学索引; 世界卫生组织; 西太区

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2016.08.018

Analysis of Data Submission Ability of Journals Collection by WHO Western Pacific Region Index Medicus FANG An, YANG Chen-liu, WANG Jun-hui, Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 The paper introduces WHO Western Pacific Region Index Medicus (WPRIM), current ways of data collection by WPRIM and existing problems. Based on questionnaires, it analyzes Chinese journals in WPRIM from the perspective of the data submission ability of journals, inclined submission ways and work suggestions. Also, it discusses problems reflected by the survey results, including the improvement of survey methods, data completeness of WPRIM and data submission ability of the editorial department, and the promotion of WPRIM publicity.

〔Keywords〕 Western Pacific Region Index Medicus (WPRIM); Global Health Library (GHL); Global Medical Index (GMI); WHO; Western pacific region

1 引言

为了促进全球健康信息的共享与利用, 世界卫生组织 (World Health Organization, WHO) 2005 年启动了知识获取项目的一部分——全球卫生图书馆

(Global Health Library, GHL) 项目, 使需要健康信息的人能够获得可靠的健康信息^[1-2]。GHL 项目是一个基于互联网的虚拟图书馆, 由 WHO 成员国共同参与, 其一项重要工作是建立全球医学索引 (Global Index Medicus, GIM), 提供全世界的医学文献题录及文摘。WHO 将成员国分为非洲区、美洲区、东地中海区、欧洲区、东南亚区和西太区 6 个区, 每个区分别建立各自的医学索引。WHO 西太平洋地区医学索引 (The Western Pacific Region Index Medicus, WPRIM) 是 GIM 的重要组成部分, 其

〔修回日期〕 2016-04-20

〔作者简介〕 方安, 副研究馆员, 发表论文 30 余篇; 杨晨柳, 实习研究员; 王军辉, 助理研究员。

目标是建立网上医学索引系统, 收集 WHO 西太平洋各成员国和地区出版的覆盖卫生、生物医学领域的期刊题录 (包括文摘) 信息, 实现西太区医疗卫生领域相关信息的全球共享, 促进世界范围内卫生相关信息的传播利用。作为 WHO 合作中心和 GHIL 中国委员会主任单位, 中国医学科学院医学信息研究所 (以下简称信息所) 受 WHO 西太区办事处委托成立了专门的 WPRIM 建设工作组 (以下简称工作组), 负责 WPRIM 检索服务平台的开发工作。WPRIM 检索服务平台于 2010 年 5 月正式上线^[3-4], 截至 2016 年 1 月 18 日, WPRIM 数据库收录来自 12 个西太区成员国出版的 594 种期刊所发表的 588 538 篇题录数据, 其中约 40% 同时提供全文链接地址。

WPRIM 当前数据采集主要通过以下 3 种途径获得^[5]: (1) 在线提交——各国数据管理人员通过账号远程登录系统后台, 按照系统流程手工填写或批量导入 XML 文件。(2) 网络采集——针对具有成熟的国家级医学文献检索系统的韩国、日本和中国 3 个国家, 工作组依据达成的相关协议通过网络途径获得数据。(3) 邮件传递——工作组通过邮箱定期或不定期接收发展相对落后国家的管理人员发来的数据。上述数据采集方式是在 WPRIM 项目初期根据不同国家的实际情况而分别采取的最优方案, 到目前为止, 均已相对固定和成熟; 但统计分析数据发现, 当前 WPRIM 数据库仍存在部分数据质量不高、部分入选期刊论文数据缺失、数据更新不及时等问题, 亟待寻求更优化、更合理的数据采集途径。期刊编辑部作为论文数据的原创者, 可以在第一时间收集本期刊最新出版的论文题录数据, 因此, 由 WPRIM 入选期刊的编辑部直接向 WPRIM 数据库提交数据可以作为当前 WPRIM 数据采集方式的有效补充。

为了解 WPRIM 入选期刊在向 WPRIM 提交数据过程中的问题、诉求以及自身能力, 工作组于 2015 年 11 月针对中国入选期刊开展了专项调查工作。本次调查主要以问卷调查方式进行, 通过电子邮箱向当前 WPRIM 收录的所有中国期刊的编辑部发放问卷, 主要内容涉及期刊基本信息、自身数据提交

能力、对数据提交方式的倾向性以及和信息所的工作建议几方面。本文主要分析本次调查的具体结果, 并针对调查结果所反映的问题进行讨论。

2 WPRIM 中国期刊数据提交能力调查结果分析

2.1 问卷回收统计

截至 2015 年 11 月, WPRIM 共收录中国期刊 276 种。本次问卷调查针对全部 276 种期刊的编辑部分别发放了调查问卷, 提交反馈调查表的期刊 (以下简称反馈期刊) 共 88 种, 所占比例约为 31.9%。各年度 WPRIM 收录中国期刊数量和反馈期刊数量, 见图 1。

2.2 数据完整性

WPRIM 目前只收录含有英文题录的期刊论文数据, 因此即使 WPRIM 收录了某一种期刊, 也只是收录其中的英文题录信息。当前针对中国入选期刊的英文题录数据, 工作组除了定期从 SinoMed 采集外^[6], 还根据协议批量接受 GHIL 中国委员会成员单位一万方数据的更新数据^[7]。根据反馈期刊编辑部提交的数据统计结果, 所有反馈期刊在入选 WPRIM 后出版的英文题录文章总数为 154 519 篇, 其中已经被 WPRIM 数据库中所收录的文章数为 93 345, 占比约为 60.4%。各年度反馈期刊在入选 WPRIM 之后所出版的英文题录文章数、WPRIM 数据库已收录的英文题录文章数以及在万方数据中的英文题录文章数的比较, 见图 2、图 3。

分析可知: (1) WPRIM 数据库中当前收录的英文题录数量与收录期刊出版的全部英文题录数量之间还有一定差距, 亟需进一步提高文献数据的完整性, 以更好地促进西太区医疗卫生领域相关信息的传播和共享。(2) 万方数据所提供的英文题录数据在 WPRIM 数据库中占有较大比例, 是 WPRIM 数据的重要来源和保障。(3) 新的数据采集途径可作为 WPRIM 题录数据来源的有益补充, 如定期采集 SinoMed 数据、直接接受来自编辑部的数据等。

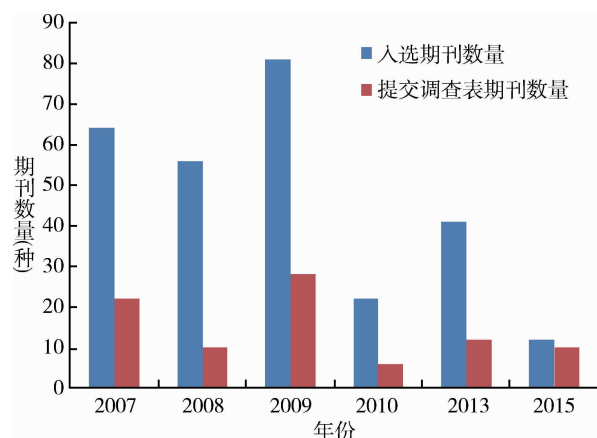


图 1 各年度 WPRIM 收录中国期刊数量和反馈期刊数量

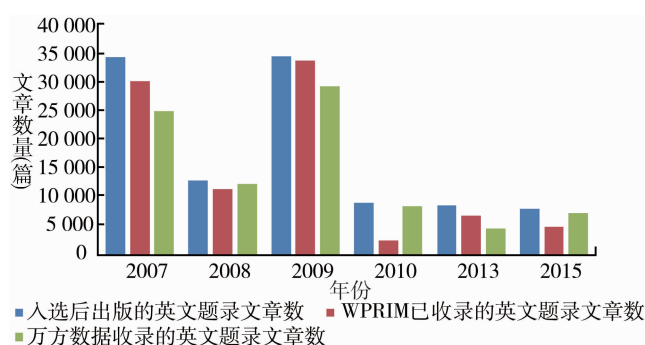


图 2 各年度反馈期刊入选 WPRIM 之后出版的英文题录文章数

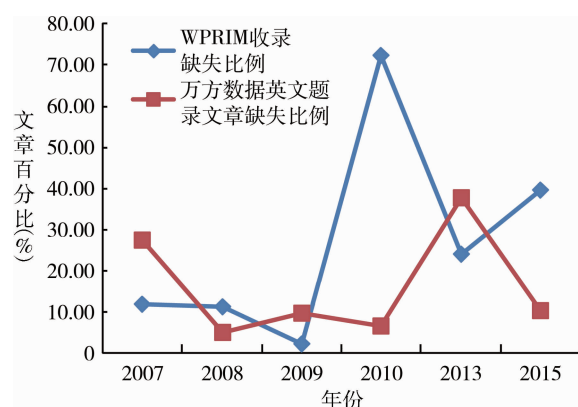


图 3 WPRIM 和万方数据缺失的各年度反馈期刊英文题录文章比例

注：WPRIM 收录缺失比例 = (入选后出版的英文题录文章数 - WPRIM 已收录的英文题录文章数) / 入选后出版的英文题录文章数；万方数据英文题录文章缺失比例 = (入选后出版的英文题录文章数 - 万方数据收录的英文题录文章数) / 入选后出版的英文题录文章数。

2.3 数据提交能力

可扩展标记语言 (Extensible Markup Language, XML) 是一种用于标记电子文件使其具有结构性的标记语言, 是网络环境中传递数据的国际通用标准^[8]。期刊编辑部的数据提交能力在很大程度上取决于其制作 XML 文件的能力。调查发现在反馈期刊中仅有 12 种期刊的编辑部表示有能力独立完成 XML 文件制作, 而 74 种期刊的编辑部则表示目前还没有能力制作 XML 文件, 需要委托相关公司才可以完成, 详细比例, 见图 4。

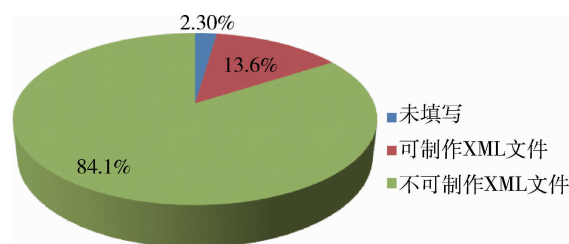


图 4 反馈期刊中不同 XML 文件制作能力的期刊比例

2.4 数据提交方式倾向性

调查问卷提供了 5 种数据提交方式供多项选择, 分别为: (1) 编辑部人员登录 WPRIM 在线数据管理系统, 手工填写表单逐条录入论文英文题录数据。(2) 编辑部本地制作符合 WPRIM 元数据标准的 XML 文件, 登录 WPRIM 在线数据管理系统后, 批量导入至 WPRIM 数据库。(3) 编辑部从被收录的其他数据库中导出自身英文题录数据, 本地转换成符合 WPRIM 元数据标准的 XML 文件后, 批量导入至 WPRIM 数据库。(4) 编辑部委托信息所加工数据并支付加工成本费用。(5) 其他方式 (由编辑部填写)。基于调查问卷的结果统计, 见图 5。结果显示, 委托信息所加工数据并支付加工成本费用是反馈期刊编辑部最倾向的方式, 选择该项的期刊数为 32, 占总反馈期刊数的 36.40%; 而编辑部希望由自己来主动提交数据的积极性也很高, 前 3 个选项的比例总计达到了 42.10%。通过对选项 5 的分析发现, 除了备选的 4 种方式外, 少部分编辑部倾向于以下 3 种方式: (1) 寄送杂志至 WPRIM

数据库。(2) WPRIM 自行采集并加工数据。(3) WPRIM 提供相关培训, 组织编辑部人员学习如何加工提交数据。

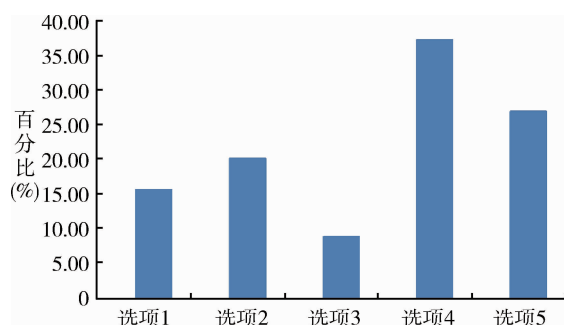


图 5 反馈期刊针对各种数据提交方式的选择比例

2.5 工作建议

调查问卷提供了 5 种建议选项供多项选择, 分别为:(1) 组织国内技术培训 (XML 制作、WPRIM 在线数据管理系统使用等)。(2) 举办编辑相关学术论坛 (英文编辑出版、国际影响力的等方面)。(3) 组团参加亚太医学期刊编辑协会 (Asia Pacific Association of Medical Journal Editors, APAME) 等年度国际会议^[9]。(4) 进一步推进 WPRIM 的宣传推广工作。(5) 其他 (由编辑部填写)。基于调查问卷的结果统计, 见图 6。结果显示, 反馈期刊编辑部对参加 WPRIM 技术培训、相关学术论坛及国际会议的积极性均很高, 其中组织 WPRIM 技术培训最受欢迎, 选择比例为 80.70%。同时, 6.80% 的反馈期刊编辑部提出了一些其他建议, 主要包括:(1) 颁发 WPRIM 收录证书。(2) 组织关于 WPRIM 遴选标准及申请流程的培训。

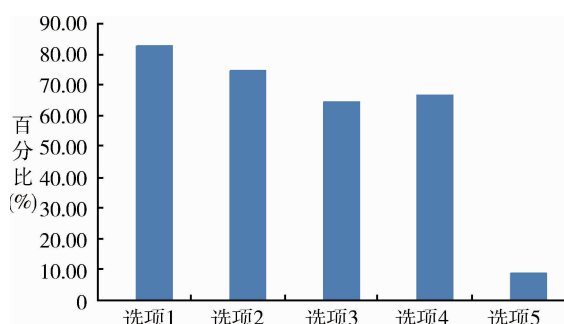


图 6 反馈期刊针对工作建议的选择比例

3 讨论

3.1 改进调查方式

本次调查是通过电子邮箱的方式将调查问卷发送给入选 WPRIM 的 276 种中国期刊的编辑部, 编辑部填写调查问卷后再通过邮箱返回给工作组。88 种反馈期刊数和入选的中国期刊总数的差距很大, 因此获得的各项统计数据均存在一定的局限性。后续可以尝试采用在线调查问卷的方式, 辅以电话通知, 以避免单纯依赖邮箱传递的片面性, 来提高调查问卷的召回率。另外, 本次调查仅针对中国的入选期刊, 尚未涉及其他国家的期刊, 而面向入选 WPRIM 的全部 594 种期刊进行调查分析, 将可以获得更全面、更客观的数据。

3.2 提高 WPRIM 数据完整性

在 WPRIM 筹备初期曾明确提出将尚未被国际知名检索系统 (如 PubMed) 收录的西太区优秀期刊文献收集起来, 建立西太区医学索引的目标^[10]; 但项目在后续推进过程中, 出于区域医学信息完整性的考虑, 部分已经被 PubMed 收录的期刊也准予加入。然而, 与整个 WHO 西太区所出版的生物医学期刊相比, 当前 WPRIM 所收录的期刊仍只是一小部分, 尤其是中国和日本, 许多相对质量较好的期刊尚未被 WPRIM 收录。而且, 从上述基于中国期刊的调查分析可见, 当前 WPRIM 数据库中的论文数据尚未能很好地覆盖已收录中国期刊实际出版的论文量, 需要扩展论文数据来源途径, 而不仅仅限于针对已有数据库的批量采集。因此, 当前除了要继续推进每年一度的期刊评审推荐工作外, 更重要的是提高已入选期刊所发表的论文数据的完整性。

3.3 提升编辑部数据提交能力

通过对反馈期刊数据提交能力及倾向方式的分析可知, 一方面, 编辑部希望由自己来主动向 WPRIM 数据库提交数据的积极性很高; 另一方面, 又由于自身 XML 文件制作能力的缺乏, 许多编辑

部不得不选择通过支付加工费用的方式委托信息所进行本期刊发表论文数据的加工。而超过 80% 的反馈期刊对举办 WPRIM 技术培训班的建议,更进一步表明期刊编辑部主动提交数据的愿望。因此,工作组除了要着手准备接受部分期刊委托进行论文数据加工的业务工作外,还应尽快组织国内 WPRIM 中国期刊的技术培训班,针对编辑部工作人员进行 XML 制作和 WPRIM 在线数据管理系统使用的培训,提升编辑部的数据提交能力。从 2011 年开始,工作组共举办了 3 次数据库管理国际培训班,以指导参与 WPRIM 建设的各国图书馆员和技术人员完成本国 WPRIM 数据的收集及提交工作,并取得了不错的效果。然而,由于培训班规模限制及资金支持问题,上述培训班主要针对的是除中、日、韩以外的其他 WPRIM 参与国家,并未涉及国内入选期刊编辑部。因此,后续应逐步考虑专门针对国内 276 种入选期刊编辑部举办 WPRIM 数据管理的培训班。

3.4 推进 WPRIM 宣传工作

到目前为止,尽管 WPRIM 项目取得了显著进展,但无论是在全球层面还是西太区内部,其影响力均有待进一步提升^[5]。仅就面向的用户来说,医药卫生领域的研究和工作人员、期刊编辑工作者及公众对 WPRIM 项目了解和服务的应用也相对有限。本次调查中,超过 60% 的反馈期刊也认为 WPRIM 应进一步推进自身的宣传推广工作。因此,工作组理应大力推进 WPRIM 的宣传推广工作,除了可以在国内举办编辑出版相关的学术论坛外,还应积极组织国内期刊编辑部参加 WPRIM 相关的国际会议等。

4 结语

为了更好地推进 WPRIM 中国期刊数据采集工作,了解 WPRIM 入选期刊在提交数据过程中的问题、诉求以及自身能力,本次调查工作主要从期刊自身数据提交能力、数据提交方式倾向性及工作建

议等方面向所有 WPRIM 中国期刊进行调查。尽管反馈期刊有限,但经过统计分析,仍发现一些宝贵的信息和良好的建议,为工作组下一阶段的相关工作指明了方向。后续,在推进相关工作的同时,还应针对此次调查方式的不足,扩大调查范围,面向所有 WPRIM 参与国家的所有入选期刊进行调查分析,以获得更合理的调查结果,指导 WPRIM 项目的建设和发展。

参考文献

- 1 World Health Organization. Global Health Library [DB/OL]. [2016-01-20]. <http://www.globalhealthlibrary.net>.
- 2 钱庆,郑英,代涛. 全球卫生图书馆与西太区医学索引项目在中国的发展概况 [J]. 医学信息学杂志, 2008, 29 (5): 1-6.
- 3 WHO Western Pacific Regional Office. Western Pacific Region Index Medicus [DB/OL]. [2015-07-31]. <http://www.wprim.org>.
- 4 钱庆,方安,代涛,等. 西太平洋地区医学索引设计与实施 [J]. 现代图书情报技术, 2010, 26 (3): 71-75.
- 5 王军辉,钱庆,方安,等. WHO 西太平洋地区医学索引建设进展与问题 [J]. 中华医学图书情报杂志, 2014, 2 (2): 75-79.
- 6 中国医学科学院医学信息研究所/图书馆. 中国生物医学文献服务系统 [DB/OL]. [2016-01-20]. <http://sinomed.imicams.ac.cn>.
- 7 北京万方数据股份有限公司. 万方数据知识服务平台 [DB/OL]. [2016-01-20]. <http://wanfangdata.com.cn/>.
- 8 The World Wide Web Consortium. Extensible Markup Language (XML) [EB/OL]. [2016-01-20]. <https://www.w3.org/XML/>.
- 9 WHO Western Pacific Regional Office. The Asia Pacific Association of Medical Journal Editors [EB/OL]. [2016-01-20]. <http://www.wpro.who.int/apame/en/>.
- 10 WHO Western Pacific Regional Office. Regional Workshop of National Focal Point Librarians [EB/OL]. [2016-01-20]. http://wprim.whocc.org.cn/news/201009/t20100902_75.htm.