

电子期刊使用数据的统计及标准化研究

□ 蔡菁 / 深圳图书馆 深圳 518036

摘要: 电子期刊作为图书馆数字馆藏的重要组成部分, 反映其利用状况的使用数据 (Usage Data) 是成本绩效测评的重要数据来源。文章从阐述电子期刊使用数据的价值入手, 进而剖析电子期刊使用数据标准化的原因, 并通过介绍美国、欧洲等若干国际性组织和机构的代表性项目来详述电子期刊使用数据标准化研究的进展, 最后对图书馆统计电子期刊使用数据的各种途径和利弊进行了分析。

关键词: 电子期刊, 使用数据, 标准化

DOI: 10.3772/j.issn.1673—2286.2012.07.010

近年来, 学术图书馆经费的有限性和数字馆藏尤其是期刊数据库价格的昂贵, 使得馆藏资源的成本绩效评价一直是图书馆颇为重视的成本控制环节。期刊资源作为图书馆馆藏的重要组成部分, 在其成本绩效评价阶段中, 反映期刊利用状况的使用数据 (Usage Data) 是必不可少的测评数据, 与成本数据相结合共同为成本绩效测评提供重要的数据来源。

在传统印刷型期刊为主体的馆藏背景之下, 引用数据 (citation data), 例如 ISI 的期刊引用报告 (*Journal Citation Reports*) 和当地期刊利用报告 (*Local Journal Utilization Reports*) 所提供的期刊引用数据, 被学术图书馆馆员视为是否保留或取消某种期刊订购的重要参数^[1]。但引用数据并不能完全反映期刊的使用状况, 譬如很多

本科生和研究人员也阅读了大量期刊文献, 但并未发表相应的学术论著, 仅是用于学习、教学或其他研究目的^[2]。此种情况下, 由于缺少精确的期刊利用数据, 增加了馆员制定馆藏订购和管理决策的难度; 同时, 用于研究印刷型期刊使用状况的传统方法也具有一定的缺陷, 不能收集到全面的各类型的期刊使用数据。

随着万维网的发展和信息技术的革新, 越来越多的图书馆开始订购并提供电子期刊资源的使用, 出版商和图书馆能够通过网络平台统计电子期刊及数据库的使用数据, 这对于出版商和图书馆来说, 是一场重要的变革。通过使用商业软件或自己开发的程序对网络日志进行数据统计和分析, 图书馆和数据库商能够较为全面地获取到有关电子期刊利用状况的使用数据信息。

1 电子期刊使用数据的价值

尽管电子期刊使用数据在学术图书馆的应用还未广泛化和标准化, 但它对于学术图书馆来说具有非常重要的价值, 主要体现在以下三个方面。第一, 使用数据能够精确地反映电子期刊馆藏的利用状况, 为电子期刊经费预算的增加提供合理的支撑依据^[3]。第二, 使用数据展示了用户如何使用图书馆的数字资源, 以及哪些资源的使用率最高; 这些数据和图书馆网站的总体利用状况等, 为重新构建图书馆网络门户、需向用户提供哪些培训课程、以及需重点突出哪些数字资源产品提供了重要的参考信息。第三, 使用数据能辅助图书馆员制定有关电子期刊购买和管理方面的馆藏决策。在购买新的电子期刊或者

在已订购电子期刊的使用合同到期之时,是否需继续订购成为图书馆需要考虑的关键问题。在处理这样的棘手问题时,将电子期刊使用数据纳入进来作为衡量和决策的依据,使决策的制定更加科学化、合理化。图书馆可通过使用统计分析来调整采购决策、优化现有资源;并可预测并发用户的增减,划定回溯数据的购买起始点、选择永久保存的电子期刊品种^[4]。

由于电子资源产品涉及不同类型,使用数据尚未完全标准化和各种来源的使用数据的不可比性,以及将电子资源的使用率及其价值相关联的难度,电子期刊使用数据在图书馆馆藏决策制定和成本绩效测评的应用方面还未达到非常广泛的程度,对于一些图书馆来说还存在一定的困难。

2 电子期刊使用数据标准化的原因

电子资源使用统计,是某个单位或组织,利用一定的方法和手段对某种电子资源在某段时间内的用户访问、检索和下载数据等情况进行科学的统计^[5]。而出版商或数据库商提供的电子期刊使用统计数据需进行标准化的最主要原因,是当前的情况下不同出版商或数据库商所提供的使用数据报告的测度标准不同,不具有可比性。

使用数据报告里虽然指标一致,比如都包含“检索次数”(number of searches)等数据指标,但最终的计算结果却五花八门。例如,一个出版商所提供的检索次数可能是从整个产品本身出发,而另一个数据库商可能是从其捆绑包的每一个特定的数据库出

发来计算检索次数。同时,登录数(log-ins)和任务数(sessions)的计算方法也不够标准化。一次任务(session)是指用户通过某一IP地址登陆网站对一个数据库进行一次成功的请求,从用户使用数据库连接成功到任务完成后退出或超时为止的过程,其中的“超时”(time-out)时间也不统一,有的系统会设为7分钟,有的系统则设为30分钟。这样,数据库商在计算各项使用数据指标时,不同的超时时间就会影响到任务数和登录数的计算以及任务的进程长短。数据库商也一般不会提供如何计算各项数据指标的具体信息或报告,这更是增加了使用数据标准化的难度。此外,数据库商所出具的电子期刊使用数据报告缺乏相应的监督机制,质量不一,通常代表了数据库商的自身利益,不能完全真实地反映电子期刊的使用数据状况。

由于出版商和数据库商以不同的方式来收集和定义使用统计数据,但均使用统一的术语——任务数、检索数、全文下载数等等,对于图书馆员来说,精确地区分和比较不同数据库商所提供的使用统计数据,难度较大;且数据报告存在质量不一和缺乏监督管理的机制,不能够真实反映电子期刊的使用绩效状况。但鉴于馆藏管理的需求,图书馆员又须对各项电子期刊产品的使用状况进行精确的比较和分析。此种情形之下,使用数据标准化的需求促使数据库商和图书馆达成一致,共同就电子期刊使用数据各项指标的定义和收集方式进行一系列标准化工作。

3 电子期刊使用数据标准化研究的进展

3.1 JSTOR项目

电子资源(包括电子期刊)使用统计数据收集的标准化工作开始于20世纪90年代末期,最早是由JSTOR网络统计工作小组主动制定了有关网络资源使用测度的一套核心评价指标和指南^[6]。该项目小组重点旨在识别电子资源使用数据评价指标的核心单元,扩展数据库商和出版商在电子资源使用评价方面的能力,并提出了使用统计数据评价和分析的模式及报告格式^[7]。该项目小组最早为数据库商制定了有关电子资源使用统计数据类型的指南,同时,在提供给图书馆的电子资源使用数据报告的质量方面,JSTOR也始终处于领先的地位。

3.2 ICOLC指南

国际图书馆联盟集团(ICOLC, International Coalition of Library Consortia)是一个在电子资源使用数据统计指南制定方面扮演着关键角色的机构。该集团创建于1996年,是一个由来自世界各国的图书馆联盟组成的非正式自发性组织。该集团主要服务于高等教育机构,其宗旨在于“为联盟的共同利益服务,促进联盟之间的对话交流”^[8]。1998年11月,ICOLC在JSTOR网络统计工作小组所制定指南的基础上制定了“基于Web的索引、文摘、全文资源的使用统计评价指标指南”(Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full-Text Resources)^[9]。ICOLC指南(后于2001年12月修改)列出了出版商数据报告中应提供的必需的数据类

型元素。而且,该份指南还包含了有关隐私和用户隐私话题、图书馆和联盟隐私权的指南,同时提供了大量的数据统计资料,以及如何利用这些报告中的数据。可以说,ICOLC提供的数据统计指南成为当时出版商用来比较数字资源使用数据的标准。但在随后的实践中,尽管一些出版商声明提供符合ICOLC标准的使用数据报告,但还是有很多出版商的数据报告并不遵从ICOLC的标准。2001年,由黛博拉·布莱希克(Deborah D. Blech)等人对包括出版商、数据库商和图书馆联盟在内所提供的电子资源使用数据报告进行了综合性的详细分析,作者发现^[10]:在2000年秋季所调查的51个数据库商所提供的数据报告中,只有3份遵守ICOIC标准,28份仅是部分采用ICOIC的数据标准,剩余20份完全不符合标准。因此,就当时来说,电子期刊和数据库等电子资源的使用数据标准化仍是一个亟待解决的问题。

3.3 ARL E-Metrics项目

1999年10月,ARL统计和测度委员会和ARL研究图书馆和管理委员会共同发起了ARL新测度标准初始计划,共包括五个项目。其中的一个项目E-Metrics(电子资源评价标准)研究,用来探究有效表述和测度电子资源使用以及价值的各种数据类型。从2000年5月到2001年12月,E-Metrics项目由佛罗里达州立大学的信息利用和管理政策研究所来承担,并得到了24个ARL成员馆的资助。E-Metrics项目共包括三个阶段^[11]:第一阶段,描述了参与研究的图书馆如何收集和使用电子

资源(包括数据库、电子期刊、电子书,还有图书馆网站和网络服务)使用统计数据的现状,同时描述了有效的数据统计标准和评价指标。第二阶段,制定和分析数据统计标准及绩效评价指标,并针对参与项目的图书馆进行指标的范围测试,并推荐使用。第三阶段,研究人员创建了培训模块和相应的工具,以帮助图书馆将数据统计方法和绩效测评标准制度化。同时,该项目还制定了有关数据统计的指南手册。E-Metrics项目的目标是就数据库使用数据的标准化问题进行商讨,同时提出一种维持电子资源使用统计和绩效测评标准发展的外部支持计划。E-Metrics报告和所制定的数据统计指南对于每个从事统计或利用电子期刊使用数据的馆员来说,都是被高度推荐的参考指南。虽然ARL项目的研究范围不仅仅局限于电子资源使用统计测度,但它所提供的有关电子资源使用数据测度标准的研究,为电子资源使用数据的标准化研究工作做出了重要贡献。

3.4 CLIR白皮书

另一份有关电子资源使用数据统计的重要性文件是图书馆和信息资源委员会(Council on Library and Information Resources,简称CLIR)于2000年发表电子期刊使用数据统计白皮书(White Paper on Electronic Journal Usage Statistics)^[12],该文件从图书馆和出版商的角度列出了有关统计和分析电子期刊使用数据的复杂性及挑战。该份文件着重于电子期刊的使用数据统计,由朱迪·卢瑟(Judy Luther)来承担该项目研究,但

Luther提出的问题也同样适用于其他网络电子资源。Luther强调了出版商和图书馆员需要共同合作来促进电子期刊使用数据统计和报告方式的规范化及标准化工作。她还指出了基于现有的电子期刊使用数据来制定馆藏决策的复杂性,因为电子期刊使用数据偶尔会脱离馆藏发展背景或者不具有可比性。该份白皮书里,Luther极力鼓励图书馆员和出版商通力合作,共同制定电子期刊使用数据统计和分析的标准化方法和模式。

3.5 NISO统计标准

有关电子期刊使用数据的标准化工作,由美国国家信息标准组织(NISO)制定的图书馆数据统计标准(ANSI/NISO Z39.7-1995)首次发布于1968年,后于1983年和1995年两次修订。2001年2月,有关图书馆绩效评价指标和数据统计标准的NISO论坛召开,NISO标准委员会被委派对原有的ANSI/NISO Z39.7-1995图书馆统计标准进行再次修订,加入有关电子资源的新的数据统计标准,包括电子资源绩效评价标准、数据库商和出版商的使用数据统计标准和报告方式^[13]。在此基础上于2002年发布了“信息服务和使用:图书馆和信息服务机构评价和统计标准——数据字典”(Information Service and Use: Metrics & statistics for Libraries and information providers – data dictionary)。

3.6 EQUINOX项目

除了美国的几个代表性项目,还有很多致力于电子资源使用数

据标准化工作的国际性机构和组织。比如,欧洲委员会(European Commission)的EQUINOX项目(欧洲图书馆绩效评估和质量管理系统项目),从1998年11月至2000年11月,该项目在欧洲委员会的资助下开展了相应的研究。EQUINOX项目共包括两方面内容^[14]:一,通过扩展电子图书馆环境下的相关绩效评价标准,制定符合当前图书馆环境的电子资源绩效评价体系;二,为图书馆管理者设计一个用于图书馆质量管理和绩效评价的工具。在最终报告中,EQUINOX项目共提出了14个绩效评价指标,这些指标不仅包涵了对电子资源使用绩效的直接评价,还体现了电子资源服务对图书馆和用户的影响测评。

3.7 COUNTER项目

出版商和图书馆解决方案委员会(The Publishing and Library Solutions Committee,简称PALS)在线使用统计工作小组在英国成立,致力于发展一套为出版商和图书馆界普遍接受的电子资源在线使用统计的实施规范,使出版商和数据库商能够为图书馆提供统一格式和规范的数据报告。该研究小组的目标是在电子资源使用的数据元素定义、产出报告格式、提交报告的方式等方面提供规范性指导。由于PALS统计工作小组代表了图书馆界和出版界双方的共同立场,从双方的共同利益出发,因此在电子资源使用统计规范化领域产生了不可替代的重要作用。

由PALS统计小组发起的国际性项目COUNTER(Counting Online Usage of NeTworked

Electronic Resources,网络电子资源在线使用统计),正式开始于2002年3月,并获得若干国际性机构和组织的支持,包括ARL、NISO和出版商集团。COUNTER项目的目标是:通过促进电子资源在线使用数据的记录和交换来服务于图书馆员、出版商和中间商(包括数据库商),使图书馆员能够对来自不同数据库商的数据统计资料进行比较^[15]。2002年11月,COUNTER实施规范(COUNTER Code of Practice)第一版(*Release 1 of the COUNTER code of Practice for Journals and Databases*)对外发布,其中包括了COUNTER项目的简介、电子资源及使用的相关术语定义等。该实施规范明确说明了出版商需遵守的数据报告内容、格式和传递方法的相关规范。COUNTER实施规范是在以往研究项目的基础上建立起来的,如ICOLC和ARL所制定的相应标准。COUNTER最具特色之处是,2004年COUNTER声明数据库商需提交电子资源数据报告供第三方审计,促使数据库商对所提供的数据报告质量负责。COUNTER还鼓励图书馆在与出版商和数据库商签订电子资源使用许可协议时,应要求对方提供遵从COUNTER规范的数据统计报告,并将此条款加入到电子资源使用许可协议中。同时,COUNTER还专门列出了“遵从COUNTER实施规范”的出版商和数据库商列表,以鼓励更多的出版商和数据库商加入进来。以上行动都是以往的研究项目中所缺少的,同时也是COUNTER能够获得广泛应用的原因。此后,COUNTER实施规范于2005年4月发布了第二

版(*Release 2 of the COUNTER code of Practice for Journals and Databases*),后又于2008年8月发布了第三版(*Release 3 of the COUNTER code of Practice for Journals and Databases*),并得到多家出版机构和组织的支持。最新的第四版(*Release 4 of the COUNTER Code of Practice for e-Resources*)于2012年4月发布^[16]。COUNTER规范是目前电子资源在线使用统计研究取得的最重要的成果,也是最为广泛应用的使用统计标准之一,国外许多数据库商如Elsevier、Blackwell、EBSCO、Nature等都宣称遵循COUNTER标准向用户提供使用数据报告。

4 图书馆统计电子期刊使用数据的途径

对于学术图书馆而言,目前电子期刊使用统计数据的来源主要有以下三种途径:(1)图书馆联盟集团购买的部分电子期刊数据库的使用统计数据可通过集团服务器来获取。(2)通过出版商或数据库商的服务器远程访问的电子期刊数据库,数据库商会定期给用户发送统计数据,或通过专用的使用统计报告网站向用户提供数据库的使用报告。每当使用统计数据更新后,就会及时通过E-mail通知用户获取。目前外文电子期刊数据库大多通过这种方式获取使用统计数据,并且越来越多的数据库商开始遵照COUNTER规范向用户定期发送使用统计数据报告。(3)图书馆自行收集电子期刊使用数据,尤其是本地镜像数据库的使用统计数据,可通过本地服务器的日志分析来获取。

4.1 由数据库商提供使用统计数据

在使用由数据库商所提供的统计数据过程中,图书馆和数据库商之间的交流和沟通起到至关重要的作用。对于图书馆员来说,在利用由数据库商提供的电子期刊使用数据的同时,还需就数据报告的质量问题给数据库商提供反馈意见,从而促进电子期刊使用数据统计的改进和标准化进程。在此过程中,图书馆员应扮演主动的沟通角色,可通过邮件或电话等简单的沟通方式,与数据库商的代表或技术服务部门联系并进行反馈,通常情况下,为进一步地提高数据报告的质量,数据库商对具有建设性的批评意见和建议都会回应或虚心采纳。此外,图书馆员还可在电子邮件组里分享他们对某一数据库商所提供的使用数据报告的感受和评价,与其他馆的馆员就电子期刊使用数据统计的相关话题交换思想和意见,或集体就共同关心的话题向出版商和数据库商提出要求。事实上,数据库商提供的电子期刊数据报告和测评标准都处于不断变化的状态中,不时会出现新的特征和新的测评标准,图书馆员也应在此趋势下不断

提出新的要求和话题。

4.2 自行统计电子期刊使用数据

除了由数据库商提供的电子期刊使用数据外,图书馆也可采取自行统计的途径。具体的统计措施包括:通过图书馆的本地服务器日志,图书馆能够收集用户如何使用图书馆网站的相关信息,还能捕获和随时跟踪某一电子期刊数据库或其他电子资源的检索用户人数。如马歇尔·布里丁(Marshall Breeding)在其“测度和实施E-use战略”^[17]一文中详细描述了图书馆自行收集电子资源使用数据的方法。通过使用网络日志分析软件或图书馆自己编写的程序软件,图书馆能够根据网络日志里的标签信息,统计访问某一电子期刊数据库的人数。同时,图书馆还可采用在数据库页面上增加计数器的方式,以方便地查验到用户的IP、网络浏览次数等指标^[18]。此外,某些图书馆所使用的图书馆集成系统(Integrated Library System)和电子期刊管理系统(Electronic Resources Management System, ERMS),也拥有电子期刊使用数

据的统计功能,能够提供电子资源访问人数的常规数据报告。

自行统计电子期刊使用数据对于图书馆来说有利有弊,利处在于电子期刊使用数据的统计不能完全依赖于出版商和数据库商,图书馆收集基本的用户使用数据存在一定的必要性,尤其是在某些特殊情况之下,譬如针对某些数据库商不提供数据报告的电子期刊,或者数据库商提供错误的使用统计报告,以及数据库商的服务器出现故障暂时不能提供完整的数据报告时。弊端主要体现在以下两个方面:首先,数据统计途径的局限性。图书馆只能通过本地服务器统计到从图书馆网站访问电子期刊数据库的人数,而当用户通过互联网的其他地址访问该电子期刊资源时,图书馆则无法获取和跟踪电子期刊的使用动态。因此,图书馆仍需依靠数据库商提供深度的数据报告来掌握和跟踪馆藏电子期刊的使用绩效状况。其次,尽管获取电子期刊使用数据对于图书馆来说不是一件困难的事情,是比较容易操作的,但逐条获取和统计电子期刊使用数据,其任务量的庞大和工作的繁琐性导致自行统计数据的工作是一个费时费力的过程。

参考文献

- [1] NISONGER T. Use of the Journal Citation Reports for Serials Management in Research Libraries: An Investigation of the Effect of Self-Citation on Journal Ranking in Library and Information Science and Genetics [J]. College and Research Libraries, 2000, 61(3): 263-275.
- [2] BLEICIC D D, FISCELLA J B, WIBERLEY S E. The Measurement of Use of Web-Based Information Resources: An Early Look at Vendor-Supplied Data [J]. College and Research Libraries, 2001, 62(5): 434-453.
- [3] MILLER R, SCHMIDT S. E-Metrics: Measures for Electronic Resources [EB/OL]. [2012-03-10]. Keynote Speech Delivered at the 4th Northumbria International Conference on Performance Measurement in Libraries and Information Services. 2001. <http://www.arl.org/stats/newmeas/emetrics/miller-schmidt.pdf>.
- [4] 张静,阎晓弟,等.目前电子资源使用统计存在的问题及解决设想——以西安交通大学图书馆为例[J].图书与情报,2008(5):82-85.
- [5] 索传军,王建朋.国外电子资源在线使用统计研究述评[J].图书馆,2006(6):43-46.
- [6] MULLINS J. Statistical Measures of Usage of Web-based Resources [J]. The Serials Librarian, 1999, 36(1/2): 207-211.
- [7] JSTOR. Web Statistics Task Force Establishes Guidelines for Electronic Resources [EB/OL]. [2012-03-10]. JSTORNews, 1998, 2(2). <http://news.jstor.org/newsPDFs/spring1998.pdf>.

- [8] International Coalition of Library Consortia [EB/OL]. [2012-03-12]. <http://www.library.yale.edu/consortia/>.
- [9] International Coalition of Library Consortia. Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full-Text Resources [EB/OL]. [2012-03-15]. <http://icolc.net/webstats.html>.
- [10] BLECH D D, FISCELLA J B, WIBERLEY S E. The Measurement of Use of Web-Based Information Resources: An early look at vendor-supplied data [J]. College and Research Libraries, 2001, 62(5): 434-453.
- [11] JULIA C B. Measures for Electronic Use: The ARL E-Metric Project [EB/OL]. [2012-04-15]. <http://www.lboro.ac.uk/departments/lis/lisu/downloads/statsinpractice-pdfs/blixrud.pdf>.
- [12] LUTHER J. White Paper on Electronic Journal Usage Statistics [M]. Washington DC: Council on Library and Information Resources, 2000.
- [13] NISO. Library Statistics Standards Committee AY (2002) [EB/OL]. [2012-04-16]. http://www.niso.org/committees/committee_ay.html.
- [14] EQUINOX Library Performance Measurement and Quality Management System Performance Indicators for Electronic Library Services [EB/OL]. [2012-04-16]. <http://www.equinox.dcu.ie/reports/pilist.html>.
- [15] Project COUNTER [EB/OL]. [2012-05-15]. <http://www.projectcounter.org/about.html>.
- [16] The COUNTER Code of Practice for e-Resources: Release 4 [EB/OL]. [2012-05-20]. <http://www.projectcounter.org/r4/COPR4.pdf>.
- [17] Breeding M. Strategies for Measuring and Implementing E-Use [R]. Library Technology Reports, 2002, 38: 1-68.
- [18] 陈晋,林立,等.图书馆电子资源访问统计的研究[J].情报探索,2009(9):125-126.

作者简介

蔡菁 (1983-), 女, 管理学博士, 深圳图书馆馆员。研究方向: 数字馆藏成本控制, 电子期刊馆藏建设。E-mail: caiqingim@gmail.com

The Research on the Statistics and Standardization of E-journal Usage Data

Cai Qing / Shenzhen Library, Shenzhen, 518036

Abstract: Being an important part of digital collection, E-journal usage data is an important data source for cost-performance measurement of library collection. This paper firstly expounds the value of E-journal usage data, then analyzes the cause of standardization for E-journal usage data. The author also elaborates the research progress of E-journal usage data standardization, and introduces representative projects of some international organizations in the USA, Europe, etc. Finally, the author also analyzes the pros and cons of different measures for libraries calculating E-journal usage data.

Keywords: E-journal, Usage data, Standardization

(收稿日期: 2012-05-01)