

关于“教育部高校图书馆事实数据库”统计难点的探讨

张惠君 张西亚

(西安交通大学, 710049)

摘要: 对“教育部高校图书馆事实数据库”有关统计工作的难点进行归纳, 分析了问题发生的原因, 认为严谨的指标体系、详尽的填报说明是做好图书馆统计工作的保障, 并提出改进和做好这一统计工作的建议。

关键词: 高校图书馆 图书馆统计工作 事实数据库

中图分类号: G250.74

文献标识码: B

文章编号: 1006-1525(2011)01-0049-05

Inquiry into statistical knots for constructing The Fact Database of the Ministry of Education in university libraries

ZHANG Hui-jun, ZHANG Xi-ya

(Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: This article sums up the statistical knots for constructing The Fact Database of the Ministry of Education, makes an analysis of the causes of the problems. The authors of this article hold that rigorous indexing system and detailed directions for filling in the form are the guarantee to better statistical work, and propose some suggestions to improve and perfect the statistical work.

Key words: university libraries; library statistics; fact database

统计工作是图书馆管理工作的重要组成部分, 它的基本任务是要真实、准确地提供可靠的图书馆统计资料, 为各级主管部门制定政策和计划、宏观调控高校图书馆工作提供依据, 也为图书馆改进工作、提升服务提供可靠的反馈信息。^[1]

科学、实用的统计指标体系是统计工作的蓝本, 而教育部高等学校图书馆情报工作指导委员会(以下简称“教育部图工委”)则是我国高校图书馆统计标准的制定者和推行者。目前我国高校图书馆使用的, 正是全国高校图工委于2005年面向全国普通高校图书馆推出的高校图书馆事实数据统计软件, 即《教育部高校图书馆事实数据库系统》(以下简称“事实数据库”)。

从1985年的《高等院校图书馆调查表》, 到1991年的《普通高校图书馆统计表》, 再到2005年版的《教育部高校图书馆事实数据库系统》, 20年间, 我国高校图书馆的统计工作几经变迁。在内容方面, 指标体系经过多次修订, 不断完善, 在填报方式方面, 从成员馆提交纸质报表, 到目前的网上填

报, 在统计数据汇总手段上, 从过去的各省图工委手工汇总, 向全国高校图工委提交纸质汇总表、到提交汇总文件软盘, 直到今天实现系统自动汇总。

自从事实数据库建立后, 教育部图工委没有再像纸质报表时代那样, 正式发布一个统计指标体系, 但事实数据库中的所有统计项已构成了当前我国高校图书馆的统计指标体系, 只不过这个指标体系隐身于一个数据库系统中, 它是事实上的统计标准。在事实数据库中, 还对许多统计项进行了注释和说明, 注释文字组成了事实数据库的填报细则。

在日常工作中, 各省高校图工委总会收到来自成员馆的咨询, 他们在使用现行的指标体系进行统计时, 存在着诸多疑问和困难。在此, 我们对其中的典型问题做一分析, 希望通过探讨, 使事实数据库更加完善, 更加实用, 以利于统计数据更加精准。

1 事实数据库常见问题^[2]

1.1 学历与学位带来的困惑

在事实数据库中, A3是关于工作人员的统计

收稿日期: 2010-10-21

作者简介: 张惠君, 女, 副研究馆员。

项,包括人数、年龄带、学历、职称等,其中,关于学历的统计指标设置如下:

A3 工作人员

...

A3.3 学历

A3.3.1 博士

A3.3.2 硕士

A3.3.3 第二学士学位

A3.3.4 本科

A3.3.5 大专

A3.3.6 大专以下

上述指标要求按学历统计出各类人员的数量,乍一看,没有问题。对于头脑清醒、认真负责、较真儿的统计人员来说,A3.3.1、A3.3.2 这里是有“疙瘩”的。

我们都知道,近些年高校图书馆有大量的员工在职接受研究生教育,他们中的大部分人接受的是非学历教育,毕业后只有学位,没有研究生学历,这样,矛盾和困惑就来了:统计博士、硕士人数时,把他们算进去吧,与上位指标“学历”(A3.3)相矛盾,因为他们没有相应的学历;不算他们,他们确实是博士,是硕士,是反映图书馆人员素质的重要指标。年年统计,了解学位与学历区别的人岁岁纠结于此,问题出在哪里呢?

1.2 文献资源购置费“寅吃卯年”如何统计

在实际运行中,相信很多图书馆都会遇到这样的情况:当年经费不够,就先欠着商家的,下一年再付;或者正好相反,当年经费比较宽裕,就把下一年的某些费用预付了。这种当年资源购置支出与实际购入资源不相符的现象,造成了文献资源购置费(B1.1)统计标准不一致,那么,究竟该按每年实际支出计算呢,还是以当年购置的资源为准,当年的资源花了多少钱就按多少统计?

1.3 区域内馆际互借借入总量与借出总量的差异

在事实数据库中,H是关于资源利用情况的统计项,其中有关馆际互借的统计指标设置如下:

H 资源利用情况

...

H2 馆际互借借入量

H3 馆际互借借出量

...

馆际互借是馆与馆之间的服务活动,有借入就有借出,入与出是相互关联,数量相当,因此,理论上,每年各省高校图书馆区域内的馆际互借借入量

(H2)总和与借出量(H3)总和应该相等,考虑到某些馆还从省外图书馆(如国家图书馆)借入图书,那么,汇总数据中,正常情况应该是借入总量略大于借出总量。但是,事实数据库中的统计结果并非如此。以陕西省为例,2008年借入总量为3600册,而借出总量为5352册,理论与实际偏差相当大,为什么呢?

1.4 数据库个数统计为何标准不一

当年购置数据库(D3)是事实数据库中统计结果比较“乱”的统计项,我们不妨再看一组陕西高校图书馆2008年的数据。

	文献资源购置及 相关费用(元)	购电子资源 (元)	当年购置数 据库(个)
图书馆 A	21302510	10717268	58
图书馆 B	3691162	727173	125
图书馆 C	3687192	1569820	22

从上述数据我们不难发现,图书馆B当年购置数据库个数明显偏高,不但比与之经费相当的图书馆C高,还远远高于资源购置费5倍于它的图书馆A。为此,我们查阅了图书馆B2005-2008年统计数据中的D3项,发现其每年购置数据库都在120个左右。针对这一现象我们推测,是否图书馆B在统计D3时采用了与众不同的统计标准?经过与该馆统计人员沟通,得到了证实。

2 问题分析与建议

2.1 学位不等同于学历

在《中华人民共和国高等教育法》第二章中,有关学历和学位的相关内容如下:

第十五条 高等教育包括学历教育和非学历教育。

第十六条 高等学历教育分为专科教育、本科教育和研究生教育。

第二十二条 国家实行学位制度。学位分为学士、硕士和博士。

目前,高校图书馆在职攻读硕士、博士学位的人比较多,通常情况下,在职申请学位属于非学历教育,申请人在获得学位后,只表明其在学术上已达到硕士(或博士)学位的学术水平,具有硕士(或博士)学位毕业研究生的同等学力(学习能力的“力”),不涉及学历,即申请人的学历并没有改变,也不能获得硕士(或博士)研究生毕业证书。^[4]

但是,事实数据库的指标设置却将学位等同于学历,把二者的下位概念混列在一起,从而将硕士、博士当作学历的下位概念,致使这部分指标在逻辑上“拧巴”,这就是“明白人”填报时困惑的原因所在。鉴于此,建议对 A3 项进行修订,把学历统计和学位统计分别列出,尽可能使图书馆统计指标体系科学、严谨。

按照上述第十六条、第二十二条的规定,在我国,学历分为专科、本科和研究生;学位分为学士、硕士和博士。因此,建议将 A3 的相关统计指标修订为:

A3 工作人员

...

A3.3 学历

A3.3.1 研究生

A3.3.2 本科

A3.3.3 大专

A3.3.4 大专以下

A3.4 学位

A3.4.1 博士

A3.4.2 硕士

A3.4.3 第二学士学位

A3.4.4 学士

...

A2 项关于馆长学历的统计指标设置,也存在这个问题,建议一并修订。

2.2 经费统计方法不一致造成结果不统一

在图书馆日常运转中,经费支出与资源购买不同步的现象时有发生,通常有两种情况:一是当年经费充裕,提前预付下一年的费用;二是经费紧张,欠商家的费用下一年再付。

针对这种经费与资源不同步的现象,在高校图书馆实际统计时,通常采取两种方法。第一种,是以资源购买情况为准,当年买了多少资源,就按资源应付费用统计;第二种,是以经费支出为准,当年实际支付多少,就按多少统计。

上述两种方法,各有各的道理,都有存在的理由,因此,建议教育部图工委明确规定一种统一的方法,以保证统计数据的可比性和持续一致性。我们认为,第一种方法能更准确地反映图书馆资源建设的现状,因而,应当作为被规定的统计方法,其计算公式为:

文献资源购置费 = 当年实付 + 上年预付 + 当年欠费 - 当年预付 - 往年累计欠费

这个公式涵盖了购置资源时有可能发生的所有付费方式,实际情况要简单些。

2.3 技术手段影响统计结果

在 1.3 中,我们关注的是区域内馆际互借流通量的现象。关于借入总量与借出总量理论与实际不相符的问题,是由于统计该指标的技术手段落后而造成的。

就目前我国高校图书馆区域馆际互借服务而言,除少数几个省市采用计算机系统管理外,大部分地区还停留在手工管理的层面。手工管理方式下的区域馆际互借服务具有这样几个特点:(1)采用纸质证件作为馆际间的通用借书证;(2)手工记录通用借书证的借出/归还;(3)手工记录馆与馆之间的互借服务。上述特点决定了馆际互借借入/借出量的统计很难精准。

对一个图书馆来说,外校读者来借了什么书,是可以精确记录下来的,可是,本校读者向其它学校借了什么书,借了几册书,不容易统计,除非在每次读者归还通用借书证时,都详细询问,这个笨办法显然不具有可持续性和可操作性。因此,在区域馆际互借服务没有实现自动化管理的地区,各馆可以准确地统计借出量,但借入量的统计存在困难。

一般情况下,很多图书馆都默认,通用借书证流通一次算作借入一册书,很明显,这与实际情况是有出入的,因为很多读者使用一次通用借书证,可能会借几册书,这就是为什么区域内馆际互借借入总量总是小于借出总量的原因。

上述问题的出现,当然是馆际互借服务管理水平不高造成的,要解决这个问题,根本的办法是实现馆际互借服务的自动化管理,以便精确地记录每一笔互借信息,包括文献借入馆/借出馆信息、被借文献信息、读者信息等,让统计工作真正成为图书馆管理工作反馈机制的一部分。

还有网上访问量(H6)的统计,也是由于技术手段的原因,从而导致统计数据缺少可比性。对于网上访问量,有的图书馆采用最传统的方法,即统计首页点击量;有的图书馆使用网上免费软件,也有技术力量较强的图书馆,自行开发流量统计软件。当然,不排除个别图书馆“捏”一个体面的数据。鉴于此,建议教育部图工委推荐一个统一的网站访问流量统计软件。Google Analytics 是个不错的选择,它是一款免费的网站流量统计工具。^[5]或者提供一个安装网站访问统计软件的服务器,供全国高校图书馆使用,这是最理想的解决办法。

2.4 关于数字资源的计量统计

2005年,教育部图工委在对事实数据库进行改版时,删除了对桌椅板凳等家当的统计,细化了数字资源的统计。由于数字资源统计比传统资源统计要复杂些,为此,教育部图工委与CALIS管理中心于2004年专门制定了《高等学校图书馆数字资源计量指南(2004年)》(以下简称“计量指南”)以及与之配套使用的《高等学校图书馆数字资源计量指南中文数据库参照表》(以下简称“中文数据库参照表”)、《高等学校图书馆数字资源计量指南外文数据库参照表》,并于2007年对《计量指南》进行了修订。^[6]

对数字资源的统计涉及到数据库个数、电子期刊种数、册数、折算册数等多个统计指标,即便有相关的统计指南,实际操作时还是需要仔细琢磨。加之统计工作一年一次,每次重拾时总得花大把的时间回忆,难免让人对此“头大”,甚至恐惧。事实上,如果方法得当,功夫下到,关于数字资源统计的大部分项目,没有想象中那么难。

根据《计量指南(2007)》,可以将数字资源分为八个类,不同类型的数据库采用不同的计量方法,详细信息见下表:

类型	需要统计的指标	统计要点
中文电子图书	数据库个数、电子图书种数、电子图书册数	包库购买每种计一册,选购每种计二册
中文电子期刊	数据库个数、电子期刊种数、电子期刊份数、电子期刊折算册数	折算时,一种一个年份换算为一册
中文二次文献数据库	数据库个数	
中文其它数据库	数据库个数	
外文电子图书	数据库个数、电子图书种数、电子图书册数	包库购买每种计一册,选购每种计二册
外文电子期刊	数据库个数、电子期刊种数、电子期刊份数、电子期刊折算册数	折算时,一种一个年份换算为二册
外文二次文献数据库	数据库个数	
外文其它数据库	数据库个数	

上述资源类型中,电子图书的统计相对容易些,无论是包库还是选购,种数基本上是清楚的。关于电子期刊的统计,由于涉及到累积量,需要换算册数,就显得比较复杂了。其中,电子期刊种数、年份的信息是统计的关键,它主要来自三个方面:订购通知或合同上提供的数据;供应商的数据库介绍;检索平台的实际数据。

2.5 填报细则和指南是统计保障

目前,国内普通高校、民办普通高校以及独立

学院共2244所,这些院校的图书馆是事实数据库的使用对象。如此众多的使用者,要保证在使用时对统计项目的理解完全一致,是不现实的,数据库个数统计就是一个典型的例子。

统计数据库个数时,有的图书馆严格按照《中文数据库参照表》执行(如1.4中的图书馆B),有的却没有;有的只统计当年新增的数据库,有的则包括了新增和续订的部分;有的只统计了电子书、电子刊之外的数据库,有的则统计了所有数据库。方法如此不统一,结果五花八门是可想而知的,它反映了一个重要问题,即与事实数据库填报工作相关的支撑文件存在不足和薄弱之处。

我们认为,对统计指标的注释和说明,与指标体系同等重要,值得我们花费笔墨加以讨论。

3 关于填报细则和指南

对于一个统计指标体系而言,统计指标的概念表达一定要简洁、明了,如果没有充分的使用说明,很容易造成统计数据的偏差,因此,应尽量对统计指标逐项解释,甚至通过举例的方法加以说明,必要时还可制定专门的说明性文件。总之,与事实数据库填报相关的支撑文字和文件,是做好图书馆统

计工作不可缺失的重要组成部分。

从近几年我们与成员馆的互动中,以及对从成员馆统计报表的审查积累中,发现下列问题在事实数据库的填报工作中比较突出。

3.1 对数据库个数统计的若干规则不严谨

(1) 填报细则和《计量指南(2007)》有相互矛盾之处。在事实数据库的填报细则中,对“D3 当年购置数据库”是这样说明的:“指除电子书、刊以外的数据库”。按字面理解,D3只统计二次文献数

数据库和其它数据库,不包括电子书、电子期刊。但是,在《计量指南(2007)》中,关于电子书和电子期刊的计量是这样规定的:“电子书以数据库个数和电子书册数为计量单位”;“电子期刊以数据库个数、电子期刊种数和份数为计量单位”。按照这个规定,电子书和电子期刊既要统计数据库个数,也要统计相应的册数、种数和份数。规则上的不一致是显而易见的,这种不一致让头脑清醒的统计人员很困惑,不知该以哪个为准;对于稀里糊涂的人来说,可能遵照的是数据库里的填报细则,也可能采用的是《计量指南(2007)》中的规定。这种情况必然造成统计方法的不统一,致使统计结果大相径庭。

(2)《计量指南(2004)》中,关于数字资源的计量有这样一句话:“数据库个数以供应商分割的最小销售单元为计量单位”,2007年修订时仍被保留。这段文字存在两个问题,一是晦涩难懂,不易把握;二是不严谨,存在漏洞。以SCI为例,它可以分年购买,难道我们在统计D3时,要把SCI每年的资源都算作一个数据库吗?这样统计显然没有意义。我们认为,这句话在《计量指南》中显得多余,可以删除。

(3)对数字资源平台的划分有待商榷。在《中文数据库参照表》中,将万方数据划分为118个子库,想必是遵照上面所说的“以供应商分割的最小销售单元为计量单位”来划分的,即便这句话本身不存在漏洞,照此分割数据库也是值得商榷的。以万方数字资源为例,这118个子库,很多都是同一种资源类型,只不过学科不同而已,在统计时仅仅因为卖家的价格体系是按学科标价,就据此划分数据库个数,并不是科学、合理的方法。我们认为,像万方数据以及类似的数字资源平台,可以按照通俗的方法划分为期刊、学位论文、会议论文、标准等,既容易理解,又便于操作。

3.2 指标体系的设计要适应更广泛的用户群体

事实数据库的统计项设计应充分考虑各种可能性,以适应各种极端现象。例如,在事实数据库中,办公费(B1.3)设置为“不能小于100”,这个设定不够客观。我们知道,现在民办高校越来越多,很多民办高校图书馆的办公费并不划拨到图书馆,而是由学校掌握,对图书馆而言,统计时该项应该填“0”,但是,事实数据库中必须填“100”才可通过系统审核。这就让人难做了,据实填“0”吧,系统不允许;填“100”吧,显然又与事实不符。

现实中,还有更加极端的例子,有的民办高校,

一年中没有给图书馆下拨一分钱运行费,如此一来,年度经费情况(B)下属的大大小小的和经费有关的统计项,统统为0。像这种情况,系统根本不支持据实填报。

针对那些落差比较大的统计项,事实数据库应该允许个别统计项添加脚注,例如办公费,多则几十万,少则为0,那么,办公费为0的图书馆就应该加上脚注予以说明。这个属系统功能方面的问题,也期待在今后加以完善。

3.3 填报细则和《计量指南》应该尽可能详尽

不是每位统计人员都像统计指标体系的设计者那样,对每一个指标项的内涵和外延都心明如镜,因此,详细的支撑文件和文字是必须的。

日常工作中,经常有成员馆提出这样的问题:我们的设备维护费全部由学校负担,我可以填“0”吗?回答是肯定的。但是对统计人员来说,填报“0”是需要掂量的,需要有权权威部门的认可,心里才踏实。如果在设备维护费(B1.2)的原有填报细则“水电费、物业费、保洁费、维修费等在内”后增加一句:“如果上述费用全部由学校支付,则可填0”,那么B1.2的统计就十分明确了。

还有一个常被问到的问题是:我们中途换了馆长,一年中有2个馆长,该填哪个馆长呢?对于这个问题,应该在事实数据库中增加说明文字:“以年终在任馆长为准”,短短几个字,使统计人员不再有疑问。

还有些统计项,是需要举例说明的。例如,在《计量指南(2004)》中,对于电子书的界定,明确了“包括学位论文及其它类似书的出版物”,但是在《计量指南(2007)》中,却简化为“包括与书类似的出版物”,这一改动使得电子书的统计出现变数。一旦图书馆统计人员更换了,或者有人忘记了以前是怎么统计的,就有可能将学位论文遗漏,这样,既造成馆与馆之间的数据不具有可比性,也会使本馆的统计数据没有了继承性。在这个问题上,我们认为,《计量指南》应该将所有能想到的“与书类似的出版物”都列举出来,并且要穷举,这样,才可使《计量指南》真正起到指南的作用。

由于没有说明,或者注释不够通俗,而使统计结果跑偏的现象还有很多。例如,在历年对各馆统计报表的审查中,我们时常发现个别图书馆将外借书刊(H1)填报“0”,与之沟通,答曰:我们馆的书没有借给外校读者。如果将H1的注释由原来简单的“纸本”改为“本馆纸质书刊流通量”,相信就不会出现错误的统计结果了。又如,个(下转第58页)

的主题内容是一致的词同时标引造成复标。由于选择相近词义的关键词造成概念重复,使医学论文标引的几个有限关键词多了一二个无效信息点。

例 1. “定西市 18 岁以下人群 5 种微量元素检查结果分析”一文中关键词:微量元素;铁;锌;铜;钙;镁;定西。其中铁、锌、铜、钙、镁均属微量元素,关键词“微量元素”属重复标引。

例 2. “泌尿生殖道支原体感染和耐药性分析”一文中关键词:支原体;解脲脲原体;人型支原体;药物敏感试验。其中解脲脲原体、人型支原体均属于支原体,故关键词“支原体”属于复标,应删除。

当然,关键词也并不是越多越好,选择太多造成滥标,势必重点不突出,专指性差,降低了检索效果,影响了查准率。

另外,当论文报道的内容较先进,对新出现的名词,相应的主题词暂未出现,仅有上位词时,个人建议直接使用新名词,这更有助于该文献的检出,

及查新检索相关课题时减少漏检。另外,目前研究中对部分基因,研究人员更倾向于使用英文,该基因的关键词建议亦使用英文。如 survivin 基因,部分文献翻译为存活素,但用“存活素”检索时,检出的相关文献量远远少于用“survivin”,影响该文献的推广及查新检索,如 BCL-2 基因,没有中文名称,只有上位主题词,该词的关键词标引应写“BCL-2 基因”。

综上所述,论文关键词的选取在论文写作及论文编辑中是不可忽视的重要环节,遵照一定的原则选取准确的关键词,有助于文献的检出及推广。

参考文献:

- [1] 徐鸿飞, 缪宏建. 医学期刊关键词选择的多维构想[J]. 南通医学院学报, 1996, 16(4): 634-636.
- [2] 胡美君. 浅谈查新工作中如何提高查全率和查准率[J]. 浙江高校图书馆情报工作, 2005, (4): 40-41.

(编辑 朱爱瑜)

(上接第 53 页)人电脑(L5),有的图书馆只填报几台,不免让人存疑,经了解后得知,是错将个人电脑理解为“归个人使用的电脑”了,如果将 L5 原有注释“年末在用的所有个人电脑数”改为“年末在用的所有 PC 机以及笔记本电脑”,或许可以使该统计项更容易理解。

总之,任何一个统计指标体系,都离不开对统计指标进行说明的文字和文件,在制定说明性文件时,既要保持语言的准确性和学术性,还要力争使之易读、易懂。

4 结论

教育部事实数据库是我国高校图书馆统计工作的规范和标准,要使其在今后的工作中发挥更大的作用,必须做到以下三点:一是指标系统的科学化,填报说明的详尽化;二要根据图书馆事业的发展,及时更新指标体系以及相关说明;此外,对统计人员的培训,也是做好统计工作的重要环节。

参考文献:

- [1] 姜晓等. 高等学校图书馆统计[M]. 成都: 成都科技大学出版社, 1995.17.
- [2] 教育部高校图书馆事实数据库[DB/OL]. <http://162.105.140.111/calis/default.asp>.
- [3] 中华人民共和国高等教育法[EB/OL]. http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_619/200407/1311.html.
- [4] 为何在职研究生考试无学历证书只有学位证?[EB/OL]. <http://www.onjobedu.com/html/hdwd/kaoshi/50768.html>.
- [5] Google Analytics [EB/OL]. <http://www.google.cn/analytics/>.
- [6] 教育部高等学校图书馆情报工作指导委员会, 高等教育文献保障系统(CALIS)管理中心. 高等学校图书馆数字资源计量指南(2007 年)[EB/OL]. http://162.05.140.111/tugon-gwei/info/index.asp?strCode=specification_63.
- [7] 高教资源. 中国高等学校教师网[EB/OL]. <http://www.ccf.edu.cn/zhiyuan/>.
- [8] 顾健. 美国教育统计中心的大学图书馆统计[J]. 大学图书馆学报, 2008, (2): 15-21.

(编辑 朱爱瑜)