



项目编号：2018010

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称：关于高校图书馆数字资源使用统计分析系统的
比较研究

项目关键词：数字资源、利用统计、分析系统、高校图书馆

项目单位(盖章)：大连海洋大学图书馆

通信地址：辽宁省大连市沙河口区黑石礁街 52 号大连海洋
大学图书馆 邮编:116023

项目主持人：李伟

联系电话：0411-84762565

电子邮件：99211790@qq.com

提交日期：2019.05.08

项目结题验收单

1 专家验收表（**主持人所在单位**组织 3-5 名专家对项目进行验收、自评。）

项目名称	关于高校图书馆数字资源使用统计分析系统的比较研究				
主持人	李伟	职务/职称	信息咨询部主任/馆员		
所在单位	（加盖公章）大连海洋大学图书馆				
专 家 意 见	<p style="text-indent: 2em;">大连海洋大学图书馆专家小组对课题项目完成情况进行了认真检查，本课题已经按计划完成了高校图书馆数字资源使用统计分析系统的比较研究工作，对图书馆选择采用数字资源使用统计分析系统起到了重要的参考价值。</p> <p style="text-indent: 2em;">该课题项目从数据采集模式、统计报告生成方式、数据全面性和准确性、主要功能模块设置、资源访问监控五个方面对国内主流数字资源使用统计分析系统进行了比较分析；并对目前图书馆数字资源使用统计分析系统应用中存在的问题进行了研究；对图书馆数字资源使用统计分析系统的作用和采购建议做出了有价值的分析。</p> <p style="text-indent: 2em;">专家小组认真审阅了项目（编号 2018010）结题报告，听取了项目主持人的结题汇报，专家组认为该项目通过深入具体的研究和，按照计划很好地完成了预定任务，专家组认为该项目符合 CALIS 农学中心验收要求，同意项目通过验收。</p>				
专家签字					
职务/职称					

关于高校图书馆数字资源使用统计分析系统的比较研究

关键词：数字资源、利用统计、分析系统、高校图书馆

一、研究背景、目的及意义

1. 研究背景

目前，数字资源在高校图书馆文献资源建设中的比重越来越大，这一点从近年来高校图书馆采购经费的变化趋势中就可以看出，数字资源的采购经费在高校图书馆的全部文献资源建设经费中所占比重呈现逐年上升趋势，甚至有的已经超过了纸本资源的采购经费。可见，采购大量的数字资源已经成为满足高校师生教学和科研文献信息需求的重要保障。然而，随着高校图书馆数字资源量的急剧增加，如何通过有效的数字资源使用统计分析，并以此为基础切实做好高校图书馆数字资源的评价工作，以便合理利用有限经费去采购性价比高更满足读者需求的数字资源，就成为关系到图书馆文献资源建设的可持续发展和图书馆服务功能优化等方面内容的图书馆工作重心。

图书馆需要以网络建设为基础，通过资源整合、数据挖掘和策略管理等手段，解决目前存在的数字资源商不能及时准确的提供访问量报告或计量方法各异，馆内统计海量资源费时费力，数据库使用效果缺乏合理性评估，读者使用行为无从分析和管理等日益严峻的问题。

2. 研究目的

目前，国内主流的几款数字资源使用统计分析系统在功能设计上和实际部署要求上存在一定差异，本课题通过对不同产品的比较分析，提出数字资源使用统计分析系统在实际应用中必须实现的基本功能和应该实现的拓展功能，还存在哪些需要解决的问题以及该类型产品未来的发展趋势。

3. 研究意义

只有图书馆自身安装使用数字资源使用统计分析系统才能获得第一手的真实的读者数据，数据的真实性问题直接影响到图书馆的利益。通过国内几款主流图书馆数字资源使用统计分析系统产品的比较分析，得出的研究结果对现阶段高校图书馆的数字资源的采购、管理、评价有实际指导意义；帮助图书馆如何更好的利用数字资源使

用统计分析系统，如分析读者行为、优化服务方式；对图书馆选择采购应用哪种产品或者选择自行研发该系统起到重要参考价值；可以作为图书馆今后开展图书馆大数据相关工作的重要数据来源。

二、研究内容

（一）图书馆数字资源使用统计分析系统的比较分析

图书馆数字资源使用统计分析系统作为数字化图书馆建设的重要组成部分，应当具备以下功能：通过使用软件或者软件与硬件相结合的技术方案，实现在校园网范围内，全面、准确的采集读者使用数字资源的相关数据信息（登录、检索、下载等），生成遵循国际 COUNTER 标准的统计报表，并支持数字资源管理、数据深度分析、读者行为查询、资源访问监控、系统管理等主要应用功能。

本课题选择数字资源统计分析系统的用户应用层面这一角度，对目前国内几款主流图书馆数字资源使用统计分析系统产品的进行对比，从以下几方面指标分析产品的差异，存在的问题及优缺点。

1. 数据采集模式：主要有只使用软件平台以及使用软件平台和采集器硬件设备相结合两种模式。

（1）只使用软件平台的模式：通过在校网服务器端安装软件系统，配置交换机把访问 Internet 端口的全部日志数据转发到监听端口，实现对网络的监听。监控服务器通过筛选规则，存储相关日志数据。该模式在硬件配置上需要 1 台监控服务器，此种模式适合在高校无法更改网络结构的情况。

（2）软件平台与采集器硬件相结合模式：软件安装于服务器端，用于统计数据分析及 web 展现。采集器部署在网络交换机处，用于采集读者访问数字资源的相关数据情况。典型的系统全网部署应当包括一套软件和两台采集器硬件设备，其中一台采集器部署在校网出口交换机处，用于采集全校读者对外部数字资源数据库的访问情况，另一台采集器部署在图书馆中心机房交换机处，用于采集校内本地镜像资源库的访问情况。此种模式需要学校网络中心配合，采集到的数据完整全面。

分析：显然第二种方式采集的数据更全面，但是需要额外采购采集器硬件设备，如果有多校区或者多个分馆，就需要采购 2 台以上的设备，部署的难度和成本大大增加，这一点需要考虑。

2. 统计报告生成方式：是否采用 Counter 规范标准，是否支持个性化定制。

(1) 将采集到的数据按照自己制定的标准生成统计报告，没有采用 Counter 标准，需要用户对统计报告进行二次处理，不便于与其他统计报告横向比较分析；不支持个性化定制。

(2) 遵循国际标准，提供符合 Counter 标准的统计报告；支持个性化定制，如按平台、数据库、图书、期刊等提供多维度使用情况的 Counter 标准统计报告。

分析：虽然 COUNTER 规范的颁布改变了数字资源使用统计不易比较的混乱状态，但是目前仍有为数不少的出版商、服务商没有遵从 COUNTER 规范，不同商家的统计数据在类型、格式、频率、深度、计量方法及可用性等方面都存在差别，从而使不同数据服务商的统计数据缺乏兼容性。作为统一采集分析图书所有数字资源的统计分析系统是否遵循 Counter 规范就变的尤为重要了，可以说遵循 Counter 规范是一个统计分析系统最基本的要求。

3. 数据全面性和准确性：数据采集是否深入到图书、期刊级别，是否包括读者行为统计。

(1) 只采集图书馆数字资源的访问使用情况，包括登录、检索、下载等基本使用统计数据。没有关联读者信息，不提供读者行为统计。

(2) 采集数据库使用统计数据，可以细化到具体某种期刊、图书的使用数据。支持与图书馆自动化管理系统对接，获取读者数据，可实现基于 IP 或者 ID 采集读者登录后访问使用数据库的具体使用行为。

分析：支持读者行为分析才能更精准的识别不同读者和操作，实现组合条件的统计查询。可以实现多维视角的统计分析，分别从读者视角和数据库使用视角查询数据，更真实的展现读者知识点需求和趋势。

4. 主要功能模块设置：围绕统计分析，满足基本功能的模块设置和包括读者行为统计、查询功能模块设置，支持个性化功能扩展。

(1) 基本功能模块设置至少应包括：资源管理、资源使用统计、Counter 标准报告、系统管理。

(2) 支持读者行为统计，还应包括：读者行为统计、读者行为查询、资源访问监控等模块。支持其他自定义扩展功能，如监控数据实时展示。

各模块具体功能如下：

资源管理：系统支持电子期刊和电子图书管理（支持批量导入和导出功能），平台、网站和数据库管理，学科分类管理，读者分组管理。

资源使用统计：系统可以实现对数据库、电子期刊、电子图书的使用统计，以及单库的使用趋势分析。

Counter 标准报告：系统可以按平台、数据库、图书、期刊等分别提供多种类型使用情况的 Counter 标准统计报告。

系统管理：系统可以提供用户管理、权限管理、日志管理和读者分组管理功能。

读者行为统计：系统支持与学校认证系统对接，获取读者数据，可以实现读者和分组读者群对资源使用情况的统计和趋势分析。

读者行为查询：在针对读者使用情况统计的基础上，可对具体行为操作和原始访问数据进行查询。

资源访问监控：系统支持设置数据库在单位时间的使用阈值，如发生违规操作，系统将自动阻断。

分析：购买较多数据库资源，使用量较大的高校图书馆应当考虑购买功能模块设置更全面的统计分析系统。

5. 资源访问监控：是否支持设置数据库使用阈值，自动阻断违规操作。

(1) 系统不具备资源访问监控功能。

(2) 系统支持资源访问监控功能：系统支持可针对归属平台、网站、单个数据库设置监测器，自定义监测的操作类型（浏览、检索、下载）、模式（读者操作次数、资源操作总次数）、时长，并选择告警和阻断方式。当读者访问次数达到阈值时，再次访问会被系统阻断；同时，系统会向管理员邮箱发送报警邮件；并且提供违规使用数据库的日志查询。

分析：部分数据库资源有下载和使用限制，通过资源访问监控功能，可以阻止用户恶意下载，保护图书馆数字资源安全。资源访问监控功能应当是数字资源使用统计分析系统的基本功能。

(二) 目前图书馆数字资源使用统计分析系统应用中存在的问题

1. 采集数据的全面性：采用软件平台与采集器硬件相结合的数据采集模式，只能做到最大限度的全面采集校园网范围内的读者访问数字资源的相关数据情况。仍然存在几种无法采集数据的情况。

(1) 用户不在校园网范围内使用数据库资源。如部分数据库提供漫游帐号或在假期等特殊时间段内不做 IP 限制。此种情况就需要联系数据商索取相关统计报告。

(2) 不同用户使用同一帐号访问使用资源。在学生宿舍存在一名学生办理校园网使用权限, 多名学生共同使用的情况。此类情况下, 读者行为统计查询就无法精确匹配到真正使用资源的读者, 导致数据不准确, 如果存在恶意下载情况, 还需要查明正在使用资源的读者信息是否和校园网帐号信息一致。

(3) 用户通过微信端和 APP 端使用数据库资源。目前很多数据库资源都提供了微信端和 APP 客户端, 方便读者利用手持移动设备访问使用资源。微信端通常不做 IP 限制, 用户只需在微信上关注数据商提供的微信公众平台即可, 在校园网外也可以使用资源。APP 客户端通常支持在校园网范围内绑定个人帐号, 然后在校园网外就可以继续使用资源。随着提供微信端和 APP 端服务的数据库资源数量的增多, 使用这两种方式利用资源的用户也会越来越多, 此类情况产生的相关数据统计无法被数字资源使用统计分析系统采集到, 仍然需要联系数据商索取相关统计报告。

(4) 数据商提供的统计报告仍然具有参考价值。数字资源使用统计分析系统应当支持导入数据商提供的符合 Counter 标准的统计报告, 支持系统生成统计报告与数据商提供的统计报告进行数据对比, 为图书馆进行数据分析提供更全面的数据参考。

2. 用户数据的隐私性: 图书馆数字资源使用统计分析系统的数据, 包含了用户的个人信息, 如姓名、学号、学院、身份证号码、IP 地址等, 记录了用户的访问使用资源的全部数据行为, 如帐号登录、检索、下载、时长等。

此类数据涉及到了用户的个人隐私, 虽然没有这方面的法律法规, 但是图书馆对于采集何种数据、谁能分析管理数据以及如何利用数据等, 应当制定详细的规定和说明, 以免一时大意触犯个人隐私权。例如采集用户个人信息是否需要征得用户本人同意, 是否可以将收集的统计信息用于分析和识别用户个人信息。例如采集用户的个人信息用于某些特定服务, 是否需要向用户告知其权利、个人信息用途及其保护方式, 是否只有在用户知情同意的情况下才能利用基于该服务明确相关的个人信息。例如将用户个人信息公开或转给第三方, 是否需用用户知情并同意。这都是图书馆需要考虑的问题。

3. 统计数据的安全性：简单概括一下，统计数据由两部分组成，分别是数据库资源利用统计和读者行为数据统计。统计数据如何保存、保存时效、第三方共享利用等都是图书馆需要考虑的统计数据的安全性问题。

数字资源使用统计分析系统对统计数据的存储有两种方式，一种是存储在图书馆本地服务器端，一种是利用互联网服务商提供云服务存储在第三方服务器端。本地端存储图书馆可以保障统计数据的管理维护在本地，保障数据的安全性，受网络服务影响小；云服务端存储具备良好的数据恢复备份机制，数据处理效率高，但是易受网络服务影响，且数据存储在第三方，图书馆难以保障统计数据的隐私性。图书馆应当根据自身情况，采用不同存储方式的产品，无论那种方式，图书馆都应当对采集的统计数据进行安全保管，限制使用各单位做好数据备份工作，随时监督管理第三方服务的数据存储是否存在问题。

三、结论与建议

（一）图书馆数字资源使用统计分析系统的主要作用

1. 为图书馆采购数字资源提供决策依据。

图书馆数字资源使用统计分析系统提供了详细的图书馆资源访问、下载情况的使用统计报告，甚至包括具体图书、期刊等的使用统计。通过对统计数据进行分析，图书馆可以更有针对性的做好数字资源的评价工作，以便合理利用有限经费去采购性价比高更满足读者需求的数字资源，有助于图书馆文献资源建设的可持续发展。

2. 为图书馆提高文献信息服务质量提供参考依据。

图书馆数字资源使用统计分析系统可以详细记录读者的资源使用情况，还可以对读者的个人信息如操作系统类型、浏览器类型、帐户信息等重要信息进行精细化的统计，方便图书馆了解读者的行为喜好。通过对读者数据的深入分析，图书馆可以根据不同读者类型的不同行为习惯，有针对性的宣传推广图书馆数字资源、优化图书馆服务功能、提高图书馆数字资源的利用率。

3. 为图书馆开展其它业务提供数据支撑。

随着各行业对信息化建设的不断重视，各业务系统的信息化程度越来越高，图书馆数字资源使用统计分析系统可以实现与图书馆其他信息化系统对接并获取数据，比如图书馆门户网站访问使用数据、读者借阅数据、门禁系统使用数据等，共同组成图书馆大数据处理分析平台，为图书馆开展其他业务决策提供数据支撑。

（二）采用图书馆数字资源使用统计分析系统的建议

综合以上比较分析结果，图书馆应当根据需求决定选择采用具备哪些相应功能的产品。考虑到日新月异的信息化建设发展，图书馆大数据平台建设或者智慧图书馆门户建设将会是未来数字化图书馆建设的必然发展趋势。建议图书馆在采购数字资源使用统计分析系统时，优先考虑功能全面、扩展性强、易操作的产品。

四、项目成果（发表的文章、开发的软件、取得的实践效果等）

无

五、参考文献

- [1] 马建霞. 图书馆数字资源访问统计研究[J]. 图书馆杂志, 2005 年第 8 期
- [2] 高广尚. 图书馆数字资源应用统计系统的设计与实现[J]. 信息工作研究, 2013 年第 23 卷第 4 期
- [3] 张静. 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 年第 5 期
- [4] 陈陶. 图书馆电子资源利用统计与分析[J]. 图书情报工作, 2005 年 4 月
- [5] 施晓华. 高校电子资源访问管理控制系统的设计和应用[J]. 计算机应用研究, 2011 年 3 月