

第12届数字图书馆前沿问题研讨班

暨Lib 2.15会议



以电子资源访问统计分析系统 为基础构建ERMS

西南交通大学图书馆系统技术部

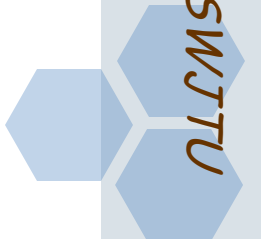
吴志强

2015年7月5日



前言

- ❖ 进行电子资源访问统计分析系统测评的原因
- ❖ 电子资源访问统计分析系统测评引发的思考
- ❖ 技术测评只是构建**ERMS**（**Electronic Resources Management System**）整体平台的切入点
- ❖ 我们希望的**ERMS**的样子



进行电子资源访问统计分析系统测评的原因

一、电子资源重要性不断增加的现状

电子资源采购量逐年增加,截至**2014**年底已占我馆图书资料经费的**56%**。



二、合理控制电子资源采购价格的需求

- ❖ 一些数据库提供商开始按照年度使用量来制定下个合同期定价等级，随着下载量的升高，下一年或下一个合同期的价格不断递增。
- ❖ 数据提供商提供的使用量统计是否合理？
比如说，由于使用多线程下载工具造成下载量虚高。
对于图书馆来说是否采信这个数据，图书馆是否有自己的数据来源以便对数据提供商构成一定程度的监管压力。

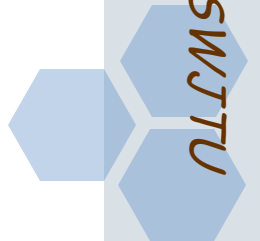




进行电子资源访问统计分析系统测评的原因

三、电子资源试用效果评价的需求

图书馆会收到各种各样的数据库、电子资源的试用申请，试用后，是否购买这些数据库和电子资源，这些电子资源在试用期间是否有足够多的读者进行过访问使用，试用效果如何？缺乏权威的定量数据支持。

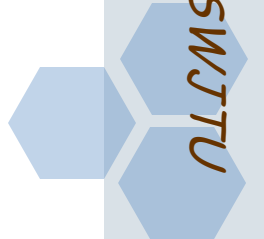




进行电子资源访问统计分析系统测评的原因

四、解决超量、违规和盗用下载的需求

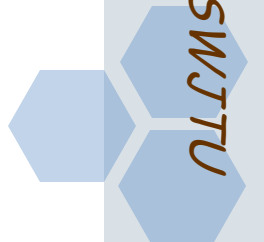
一些数据库提供商，特别是国外的数据库厂商对下载量或者
下载速度敏感度很高，经常导致图书馆收到警告邮件，
甚至
导致个别用户 I P 地址或地址段被厂商封掉。这些行为
不仅
影响了电子资源正常使用；





进行电子资源访问统计分析测评的原因

由此可见，拥有设计合理、能够满足解决以上需求的图书馆电子资源访问统计分析系统是当前图书馆更好的建设数字馆藏不可或缺的工具。

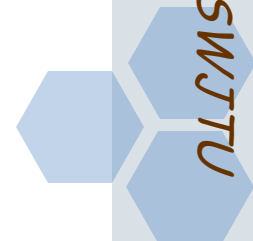




进行电子资源访问统计分析测评的原因



因此我们决定
向国内从事电
子资源访问统
计分析的厂商
发出测试邀请!





电子资源访问统计分析系统测评引发的思考

一、测试时间

- 2014年5月底开始第一个系统部署
- 2014年11月完成最后一家部署

二、参加测试的公司有四家





电子资源访问统计分析系统测评引发的思考

三、硬件条件

1) 服务器

- 每个系统都提供相同配置的联想R525两路服务器（2路4核CPU、12G内存、147G硬盘）
- 其中应某厂商要求配置2台。

2) 存储

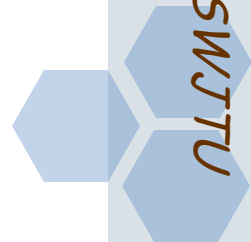
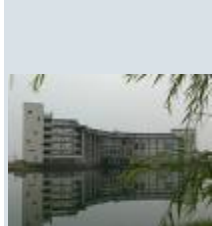
- 服务器自带147G硬盘
- 其中某厂商系统在后期通过ISCSI方式扩容500G





四、系统拓扑

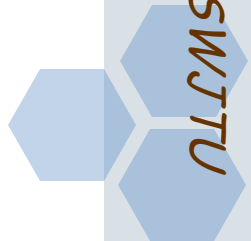
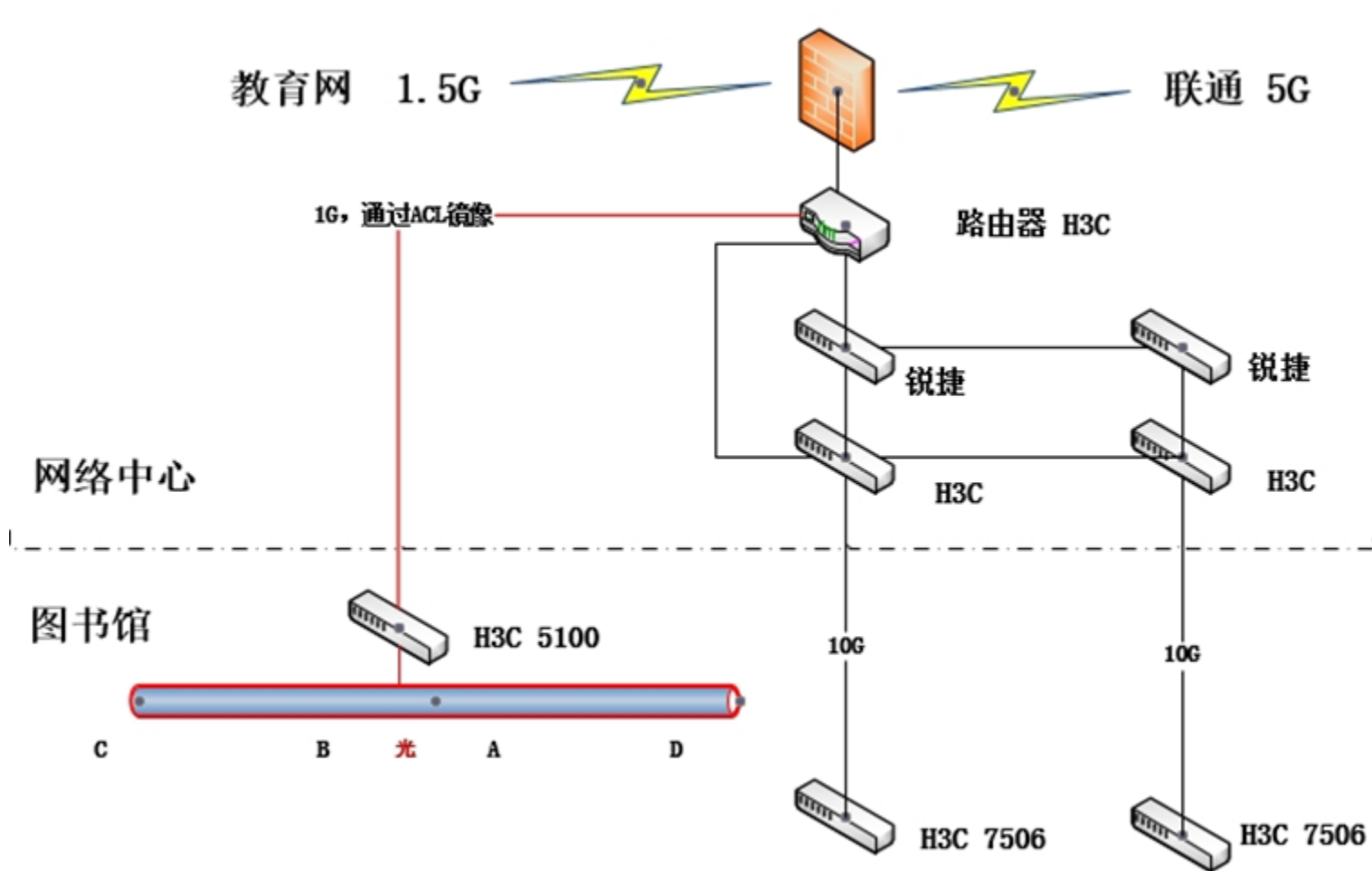
- 1) 我校成都校区有**6.5G**的出口带宽，都在九里校区网络中心
- 2) 两校区图书馆交换机与学校交换机间**10G**连接
- 3) 图书馆与网络中心间**1G**光纤连接
- 4) 在网络中心出口交换机上设置多对一端口镜像，并对网络数据包进行**ACL**过滤。
- 5) 在图书馆交换机上再次配置端口镜像，并将需要获取数据的光电端口与镜像端口划在同一**VLAN**中，在**VLAN**中广播数据包。
- 6) 丢包率小于**5%**。





电子资源访问统计分析系统测评引发的思考

测试项目总体拓扑图





五、测试方案

❖ 公平、公开是首要原则

我们没有对参评厂商的软硬件平台提出具体的要求。而是本着让参评厂商充分发挥自己产品特点，让参评厂商在我馆统一提供的相同网络环境下各自搭建自己的参评系统，充分展示自己的技术特点。

❖ 测评点的选择

通过对我们感兴趣的功能点或者是在我馆角度看来应该实现的功能点加以罗列测评，看哪些得以实现，实现的程度和效果如何。

❖ 我们只能针对我们测试时发现的厂商功能项目进行测试比较
测试时我们没发现的功能点，厂商在其后的升级中得以实现，但是我们为了保证公平，也只能以同一个时间基线加以测评。





电子资源访问统计分析系统测评引发的思考

六、测试主要分9类功能点进行：

- 1) 资源管理功能
- 2) 读者管理功能
- 3) 系统管理功能
- 4) 读者资源使用统计功能
- 5) 读者行为查询功能
- 6) 资源使用统计功能
- 7) 资源访问监控与超量访问阻断功能
- 8) Counter标准报告
- 9) 数据库IP变化的探针功能





电子资源访问统计分析系统测评引发的思考

功能测试表(局部)

硬件平台		系统平台架构	是否支持虚拟化(与商家交流)	对多校区多出口监控的解决方案
资源管理模块	资源列表	数据列表	数据列表	数据列表
	资源分类	按级	按级	按级
	资源批量导入	数据导入	数据导入	数据导入
	资源批量导出	数据导出	数据导出	数据导出
	资源手工添加	数据手工添加	数据手工添加	数据手工添加
	域名对应关系管理	域名对应关系管理	域名对应关系管理	域名对应关系管理
	多数据收集器设置	多数据收集器设置	多数据收集器设置	多数据收集器设置





电子资源访问统计分析系统测评引发的思考

功能测试表(局部)

	多数数据采集器设置					
读者管理模块	读者(院系)分组	√(分记录)	√(数据库显示结果)	√(查询分组)	√(根时查询分)	√(根时查询分)
	读者(院系)分组允许IP地址重复设置					
	读者分组可以自助修改					
	读者信息展示					
	读者信息批量导入					
	读者数据导出					
系统管理模块	用户管理					
	角色管理					
	日志管理					
	系统状况检测或运维报告-邮件					
针对读者使用资源的统计模块	按照读者分组使用资源进行统计					
	按照读者分组使用数据库进行统计	√(浏览、下载、日)	√(检索、T)	√(检索、浏览、下载)		
	按照读者分组使用资源趋势进行统计(图表方式)					
	按照读者分组使用期刊状况进行统计		√(刊名、T)	√(量、下载人统计图)		
	读者原始访问统计					
	统计数是否支持导出					
读者使用行为统计	检索					
	下载					
	浏览					
	流量					
	ip地址					





电子资源访问统计分析系统测评引发的思考

功能测试表(局部)-总共123项功能点

	平台	行为描述				
		文件名称	URL			
软件平台		数据导出	√			
		数据库检索统计	√			
		数据库浏览统计	√			
		数据库对比分析统计		√ (多)		√ (多) 显示
		数据库首次访问次数				
		数据库流量				
		数据库下载统计				
		数据库被拒量统计				
		期刊篇名下载统计				√
		期刊篇名下载人数统计				√
		期刊篇名浏览统计	√			√
		期刊篇名浏览人数统计				√
		期刊分类统计				
		期刊刊名下载人数统计		含pdf下载		√
	资源使用统计模块		期刊刊名下载统计			
		期刊平台下载统计				√ (浏览人示)
		期刊刊名浏览量统计				
		期刊刊名浏览人数统计				
		期刊被拒量统计				√





电子资源访问统计分析系统测评引发的思考

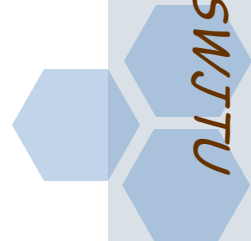
七、测试效果以及方案的选择

❖ 系统架构和功能点的选择

对于不同的馆和不同的使用场合会有不同的选择与评价。有的馆希望系统能够提供较完善的、现成的电子资源访问统计报告，并且这些报告最好符合**COUNTER**标准，而对系统技术上是如何实现这些功能的技术细节要求不多；

有的馆则恰恰相反，希望能够深入了解系统的技术细节，即系统是如何进行访问统计的。抓取数据的实时性如何？数据如何保存？元数据保存时间长度以及这些元数据如何来进行数据挖掘？

因此，即便是站在不同技术角度，对系统的测评要求差别也会很大。





电子资源访问统计分析系统测评引发的思考

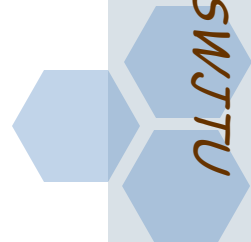
八、测评引发的思考

❖各家系统都缺乏对电子资源整个生命周期管理的完整模块;

个别厂商即便是有，也很简单。

❖电子资源访问统计分析系统应该作为ERMS系统的一个重要组成部分存在。

图书馆仅仅依靠电子资源访问统计分析系统是不够的。

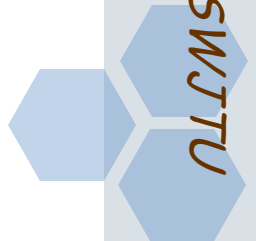




技术测评只是构建整体平台的切入点

一、通过测评发现国内厂商对ERMS重要性缺乏了解

参加测评的厂商没有认识到相对于电子资源访问统计分析系统而言，**ERMS**的技术难度不高，却是图书馆对电子资源生命周期进行管理的紧缺工具。通过电子资源访问统计分析系统采集的数据只有汇集到电子资源的整个生命周期管理中，才能发挥最大的作用。

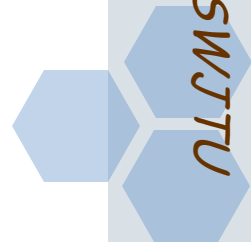




技术测评只是构建整体平台的切入点

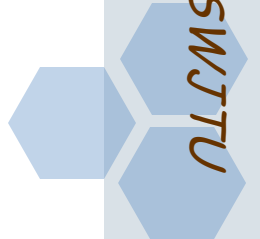
二、访问统计分析系统是对现有**ERMS**系统的重要补充

通过这次测评工作，我们发现相对现有**ERMS**系统中的数据，访问统计分析系统的数据是动态的。**ERMS**加入这个模块后，就可以以定量的形式介入整个电子资源生命周期的管理。从最初的试用数据库访问量的获得、到数据库购买开通后初期使用效果、以及稳定运行后的利用率等等，全程都能有定量的数据作为决策依据。





技术测评只是构建整体平台的切入点



三、通过技术测评完成技术路线的形成

- ❖ 在测评过程中，为了更好的了解各图书馆对访问统计分析系统和ERMS的需求情况，我馆以“四川高校图工委现代技术专业委员会”为平台，组织邀请了四川省内多所高校进行了集中测评讨论。
- ❖ 在讨论中各高校馆达成一个共识：即一个完整ERMS应该在包含访问统计分析功能的同时，汇入管理电子资源的整个生命周期过程中的所有相关资料，并能够提供接口接收各个电子资源厂商的数据报告。
- ❖ 同时我们在与各个参加测评厂商分别交流中都提出这个观点和建议，也得到了各厂商的积极响应。



我们希望的ERMS的样子

一、具有开放接口环境的ERMS

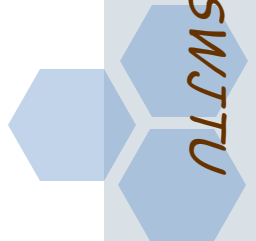
- ❖ 能够收割或者手工上传符合各种标准的统计数据，采购数据等；
- ❖ 能够对接自动化管理系统；
- ❖ 能够对接移动图书馆，以便让我们了解读者通过移动环境访问电子资源的情况；

📍 移动图书馆2014年4月正式上线

※注册用户数（2015.1.19）：**17,097**人；

※总点击量50,529,653次；平均每天点击139,584次；

※总登录量1,232,098次；平均每天登录3,422次。

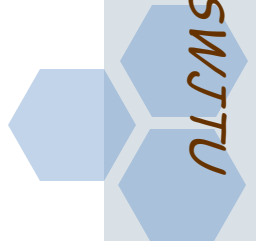




我们希望的ERMS的样子

二、具有足够厚实的访问统计数据，支持数据挖掘技术的应用

关于访问统计数据的挖掘维度与深度问题，首先要基于系统本身保存了足够多的数据，如果访问过程的原始数据已经被丢
弃，那么所谓的挖掘也就无从谈起。即便是原始数据得到了保
存，如果数据揭示方式相对固定，那么作为用户如何使用这些数据
进行挖掘也得不到启示。因此谈论数据挖掘的维度和深度，需要系统本身支持数据的多样性展示、支持或者有很好的二次开发条件。





我们希望的ERMS的样子

数据库名称	数据库种类	全文/文摘	归属网站	归属平台	检索次数	浏览次数	下载次数	开始日期	
中国学术期刊网络出版总库	中文	全文	中国知网网站	CNKI中国学术期刊全文数据库	29569	16529	25018	2014-11-21	2
中国博士学位论文全文数据库	中文	全文	中国知网网站	CNKI中国学术期刊全文数据库	28712	3018	6899	2014-11-21	2
中国优秀硕士学位论文全文数据库	中文	全文	中国知网网站	CNKI中国学术期刊全文数据库	28706	9283	12508	2014-11-21	2
学术期刊数据库	中文	全文	万方网站	万方数据：学术期刊	4217	619	2079	2014-11-21	2
学术会议数据库	中文	全文	万方网站	万方数据：学术期刊	3615	43	135	2014-11-21	2
学位论文数据库	中文	全文	万方网站	万方数据：学术期刊	2899	115	446	2014-11-21	2
Journal	外文	全文	Scimedirect主	SD(SDOS)	1715	3396	4128	2014-11-21	2
Conference Publications	外文	全文	IEEE/IETElectro	IEL(IEEE/IET Electronic Libran	1642	585	0	2014-11-21	2
Books & eBooks	外文	全文	IEEE/IETElectro	IEL(IEEE/IET Electronic Libran	1642	13	0	2014-11-21	2
Journals & Magazines	外文	全文	IEEE/IETElectro	IEL(IEEE/IET Electronic Libran	1642	1441	0	2014-11-21	2
Book	外文	全文	Scimedirect主	SD(SDOS)	1133	141	3	2014-11-21	2
Early Access Articles	外文	全文	IEEE/IETElectro	IEL(IEEE/IET Electronic Libran	492	35	0	2014-11-21	2



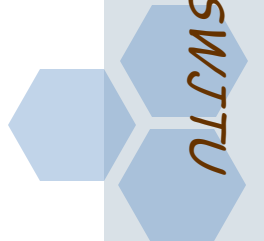
因为我们拥有足够详细的访问信息，从访问者的IP地址，到访问文献的具体篇名；从使用的检索词，到下载还是浏览的具体行为，以及正在使用的下载工具等等，理论上我们能做任何形式的数据挖掘工作，能够为特定的读者群体提供定制的学科服务。





我们希望的ERMS的样子

但是希望最后ERMS不要变成

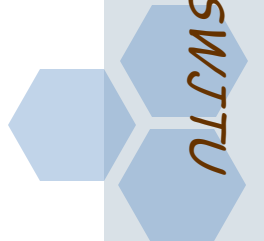




我们希望的ERMS的样子



Library of SWJTU





谢谢!

西南交通大学新校区 - 图书信息交流中心

