



COUNTER Release 5 的新特征及其应用研究

□ 杨巍 叶仁杰* 吴元业 林伟明

摘要 COUNTER Release 5 旨在应对不断变化的电子资源,提供可靠、一致和可对比的使用统计数据。文章分析了 COUNTER Release 5 的规范并探索其在中国高校图书馆数字资源采购联盟(DRAA)中的应用。研究发现,与 COUNTER Release 4 相比,Release 5 有了明显的变化,包括定义了全新的报告体系、统计指标类型、RESTful SUSHI 接口等技术细节。为在 DRAA 中应用 COUNTER Release 5,需要从两个方面改进:一是让统计模块兼容 COUNTER Release 5 的数据模型,使 COUNTER Release 5 的统计数据可以融入现有的统计模型;二是需要新增 RESTful SUSHI 接口的客户端及文件解析器,让系统可以获取并解析到 COUNTER Release 5 的报告。

关键词 COUNTER Release 5 SUSHI 使用统计

分类号 G250.7

DOI 10.16603/j.issn1002-1027.2020.01.003

1 引言

COUNTER (Counting Online Usage of Networked Electronic Resources) 是规范电子资源使用统计报告数据处理、审核和提交的国际化标准^[1],2002 年开始由高校、出版界和中间商共同发起,其目的是为在线信息服务商和用户提供可靠的、一致的、兼容的使用统计标准和方案,并获得了大部分国际著名学术出版机构的支持。COUNTER 经历了 5 个版本的升级。第一版 COUNTER 在 2003 年推出,从第一版到第三版主要关注数据库和期刊,2012 年 4 月正式出版发行了第四版《COUNTER 电子资源使用统计实施规范》,集成了在线期刊、数据库、在线图书及多媒体内容等电子资源的使用统计规范,是对原有《COUNTER 在线期刊及数据库使用统计实施规范》(第三版)及《COUNTER 在线图书和参考书使用统计实施规范》(第一版)的升级与整合,并正式统一命名为《第四版 COUNTER 电子资源实施规范》(The COUNTER Code of Practice for Release 4)(以下简称 COUNTER R4)^[2]。

最新版本的 COUNTER Release 5 于 2019 年 1 月开始生效,目前其官网已经发布了第五版《COUNTER

电子资源使用统计实施规范》(The COUNTER Code of Practice for Release 5)(以下简称 COUNTER R5)^[3],以及 COUNTER R4 与 R5 两个版本之间的过渡时间表和过渡选项,COUNTER R4 会一直服务至 2021 年 1 月,从 2019 年 1 月开始,数据库商会开始逐渐过渡到 COUNTER R5。COUNTER R5 着重于提高报告的一致性、可信度和可对比性,并在提供的报告种类、报告格式、统计量的定义、COUNTER_SUSHI 接口等细节做了很多重大的调整。目前有 40 家数据库商已经声称正在开始实现 COUNTER R5 的工作^[4],具体进度还需要继续跟踪。

目前无论是国内还是国外,对 COUNTER R5 的研究都比较少。在国外,主要是奥利弗·佩奇(Oliver Pesch)对 COUNTER R5 的新功能,与以前的实践进行了比较,并讨论了对图书馆的影响^[5-6]。在国内,胡大琴基于 R5 草案版本对比了统计指标的变化^[7];侯景丽从报告设置、指标类型、数据类型、访问类型等讨论了 R5 引入的新特征^[8]。目前国内研究主要基于对 R5 文档的解读及对应用前景的展望,较少深入分析或开展应用型研究。

* 叶仁杰,ORCID:0000-0002-2845-507X,邮箱:yrj0618@szu.edu.cn。



中国高校图书馆数字资源采购联盟(DRAA)门户于2013年初开始支持通过SUSHI(Standardized Usage Statistics Harvesting Initiative)协议自动获取COUNTER报告,于2014年开始支持COUNTER R4^[9],并向下兼容COUNTER R3。截至第16届引进数据库培训周,有329家成员馆授权DRAA为本机构收割COUNTER报告,DRAA共收集了268235个COUNTER报告,93613155条COUNTER数据量。DRAA于2015年9月正式成立资源使用统计工作组,梳理DRAA数字资源使用数据收集与统计工作的现状及问题,敦促数据库商配合使用数据收集工作,对使用数据质量进行抽查与验证,目的是建立使用数据收集的长效机制。如何及时准确地应对COUNTER R5的新特征新变化,是DRAA下一阶段统计工作面临的重大挑战。为此,本文着重分析COUNTER R5的新特征,探索其在DRAA统计模块中的应用,推动国内图书馆及供应链中的内容供应商和数据库商等对COUNTER R5标准的理解与应用。

2 COUNTER R5 的新特征

COUNTER一直致力于提供一种行为准则,帮助图书馆获得在线学术资源信息的一致、可对比、可靠的使用统计报告。COUNTER发布的实施规范定义了电子资源在使用过程中需要统计的数据量、统计数据计算的策略、输出报告的内容、格式等技术细节。随着技术的发展,电子资源的内容及其访问方式都有很大的变化,COUNTER R4中对电子资源使用统计的定义和解释已经不能满足这种变化,某些情况下还会导致实际使用环境和COUNTER定义的冲突和混淆^[10]。COUNTER R5的改变侧重于提高报告的清晰度和一致性,与R4相比,R5更新了整个报告体系,重命名了计量类型,引入了新的元素来拓展规范的适用范围,并引入了全新的SUSHI协议。COUNTER R5具有的新特征如下:

2.1 新报告体系

COUNTER R4共有24个报告^[11],每新增一种统计数据需求,就要新增一种报告类型,不同报告之间为并列关系。COUNTER R5的报告由主报告(Master Report)和标准视图(Standard View)构成,用户可以通过对主报告设置过滤参数来满足不同的统计需求,标准视图可以理解为根据常见统计需求

预设了过滤参数的主报告。COUNTER R5的这种报告体系在提供完善统计数据的同时也实现了良好的伸缩性,仅需改变请求参数即可满足不同的统计需求。

COUNTER R5包括4个主报告,分别是PR(Platform Master Report)、DR(Database Master Report)、TR(Title Master Report)、IR(Item Master Report)。4个主报告覆盖了4种主流的电子资源统计方式,包括平台、数据库、标题和项目。

标准视图是从主报告派生出来的,标准视图的报告ID格式为{Master Report_ID}_ {View ID}。COUNTER R5中共提供了12个标准视图来满足图书馆的日常统计需求(见表1)。

另外,在报告的格式方面,COUNTER R5的各类报告使用相同的报告结构,使用同样的元素、属性定义,提供表格格式和JSON格式(JavaScript Object Notation)两种不同格式的报告,并且对于同一种报告的表格格式和JSON格式数据内容也保持一致。

2.2 新属性及元素

随着数字技术的发展,出现了许多全新的电子资源载体及资源访问途径,COUNTER R5的统计范围包括期刊、数据库、电子书、文章、多媒体、研究数据以及社交媒体数据等,为了以一种一致的、可对比的方式管理不同种类资源的使用统计,COUNTER R5添加了一些新的属性和报告元素^[12]。通过使用这些属性和元素对主报告进行过滤,可以满足用户更加丰富的使用统计需求场景。新增的属性及元素主要有:

(1)主机类型(Host Types)。统计报告由不同的内容主机所提供,使用报告需求因主机类型而异,为了适应这种差异引入了主机类型的属性。但是要注意的是,主机类型属性并未出现在COUNTER报告主体中,主要是为了帮助数据库商确定不同的主机类型应与哪些统计报告、统计量相关联。主机类型主要包括以下12种:A&I_Database、Aggregated_Full_Content、Data_Repository、Discovery_Service、eBook、eBook_Collection、eJournal、Full_Content_Database、Multimedia、Multimedia_Collection、Repository、Scholarly_Collaboration_Network。

(2)数据类型(Data Types)。数据类型用来描述报告的统计内容是什么,具体有13种数据类型(见表2)。



表 1 COUNTER R5 报告体系

主报告	标准视图 ID 及名称	过滤参数	说明
Platform Master Report	PR_P1 Platform Usage	Access_Method=Regular	平台级别的使用量
Database Master Report	DR_D1 Database Search and Item Usage	Access_Method=Regular	数据库的关键词搜索、请求、调查等统计量
	DR_D2 Database Access Denied	Access_Method=Regular	访问数据库被拒绝报告,原因是访问人数超过了并发用户数,或者机构对资源没有访问权限
Title Master Report	TR_B1 Book Requests (ExcludingOA_Gold)	Data_Type=Book; Access_Type=Controlled; Access_Method=Regular	关于书籍全文活动的报告,不包括 GOA 内容
	TR_B2 Book Access Denied	Data_Type=Book; Access_Method=Regular	访问书籍被拒绝报告,原因是访问人数超过了并发用户数,或者机构对资源没有访问权限
	TR_B3 Book Usage by Access Type	Data_Type=Book; Access_Method=Regular	按访问类型(Access Type)进行分割的书籍报告
	TR_J1 Journal Requests (ExcludingOA_Gold)	Data_Type=Journal; Access_Type=Controlled; Access_Method=Regular	期刊内容的报告,不包括 GOA 内容
	TR_J2 Journal Access Denied TR_J3 Journal Usage by Access Type	Data_Type=Journal; Access_Method=Regular Data_Type=Journal; Access_Method=Regular	访问期刊被拒绝报告,原因是访问人数超过了并发用户数,或者机构对资源没有访问权限 按访问类型(Access Type)进行分割的期刊报告
	TR_J4 Journal Requests by YOP (Excluding OA_Gold)	Data_Type=Journal; Access_Type=Controlled; Access_Method=Regular	按年份的期刊访问报告,用于分析备份文件或永久使用的内容
Item Master Report	IR_A1 Journal Article Requests	Data_Type=Article; Parent_Data_Type=Journal; Access_Method=Regular	文章级别的期刊使用报告
	IR_M1 Multimedia Item Requests	Data_Type=Multimedia; Access_Method=Regular	多媒体资源项目级别的使用报告

表 2 数据类型

数据类型	描述信息	相关的报告 ID
Article	期刊或参考书中的文章。在 IR 中使用时,只有项目是 Article 才可使用	PR, IR
Book	专著	PR,DR,TR,IR
Book_Segment	图书的片段,例如一章一节等等。在 IR 中使用时,只有项目是 Article 才可使用	PR, IR
Database	统计数据为数据库级别	PR, DR
Dataset	数据集	PR, IR
Journal	在杂志或连续出版物上发表的内容	PR,DR,TR,IR
Multimedia	音频、视频、图片、流媒体音频视频等	PR,DR,IR
Newspaper_or_Newsletter	在新闻或时事通讯上发表的内容	PR,DR,TR,IR

数据类型	描述信息	相关的报告 ID
Other	不属于其他数据类型的内容	PR,DR,TR,IR
Platform	内容平台,统计数据为平台级别	PR
Report	一份报告文件	PR,DR,TR,IR
Repository_Item	机构库中的内容	PR,IR
Thesis_or_Dissertation	学位论文	PR,DR,TR,IR

(3) 章节类型(Section Types)。一些学术内容可以按章节访问,用户可以一次只访问资源的一章或一段内容。共有 5 种不同的章节类型(见表 3)。



表 3 章节类型

章节类型	描述
Article	期刊、百科全书或参考书的一篇文章
Book	作为单独文件被访问的一本完整的书
Chapter	书的一章节
Other	不属于其他章节类型的内容
Section	一组章节或文章

(4) 计量类型(Metric Types)。相比 COUNTER R4,COUNTER R5 更新了全部的计量类型,计量类型表示被统计的与内容相关的行为,主要分 3 类 12 种(见表 4):与检索相关的行为(Searches),与内容项目相关的行为(Requests and Investigations),以及与拒绝访问相关的行为(Access Denied)。其中与内容项目相关的行为分为请求(Requests)和调查(Investigations)两种。任何与内容相关的行为都被认为是调查,包括查看或下载内容等行为,而请求行为仅限于用户浏览或下载资源内容本身。

表 4 计量类型

计量类型	描述
Searches_Regular	用户对选择的数据库搜索的次数,数据库级别而非平台级别的行为
Searches_Automated	在主机站点或发现服务上执行的搜索
Searches_Federated	通过联合搜索进行的搜索行为
Searches_Platform	平台级别记录的用户搜索行为,用于平台报告
Total_Item_Investigations	内容项目或与之有关的信息被访问的次数
Unique_Item_Investigations	用户在一次会话期间对资源的访问只记录一次
Unique_Title_Investigations	用户在一次会话期间对 Title 资源的访问只记录一次,例如对期刊或电子书,支持对电子书的总体使用量进行可比较的统计
Total_Item_Requests	项目被请求的总次数
Unique_Item_Requests	用户在一次会话期间对项目进行的请求只记录一次
Unique_Title_Requests	用户在一次会话期间对 Title 资源的请求只记录一次
No_License	由于用户所在机构无权限而导致的拒绝访问次数
Limit_Exceeded	由于超出并发数而导致的拒绝访问次数

(5) 访问类型(Access Types)。为了区分 OA 资源和其他开放访问的资源与需要授权访问资源的使用情况,COUNTER R5 引入了访问类型的属性,共有 4 种访问类型(见表 5)。

表 5 访问类型

访问类型	描述
Controlled	需要授权才可以访问,另外出于商业考虑,平台提供的可以自由访问但非 GOA 的资源,也应当做 Controlled 资源
OA_Gold	用户访问 GOA 协议下的资源
OA_Delayed	延迟开放获取,目前仅收入 COUNTER R5,但是并未要求数据库商实现
Other_Free_To_Read	资源可以自由访问,但又不属于 GOA,仅用于机构库

(6) 访问方式(Access Methods)。访问方式属性用于区分正常的用户访问行为和为了数据挖掘(TDM)而进行的访问行为(见表 6)。

表 6 访问方式

访问方式	描述
Regular	用户在平台或内容主机上的行为
TDM	出于数据挖掘需求而对资源进行的访问活动

(7) 报告过滤器(Report Filter)和报告属性(Report Attribute)。COUNTER R5 允许用户通过对主报告设置报告过滤器和报告属性来定制视图。属性定义了报告中包含的统计项目(列),过滤器定义了报告中应包含的统计值。报告过滤器和报告属性应该在报告头中指明。

2.3 新 SUSHI 协议

COUNTER R5 的另一个重大变化就是定义了全新的 SUSHI 协议。COUNTER R4 使用的是基于 SOAP^[13]的 Webservice 服务,用户请求与服务器返回数据格式为 XML^[14]。COUNTER R5 使用了 RESTful^[15]接口的服务,用户请求以 URL 的形式发送给 SUSHI 服务器,返回报告数据格式为 JSON (JavaScript Object Notation)^[16]。

SOAP(Simple Object Access Protocol)是一个严格定义的信息交换协议,SOAP 使用 XML 数据格式,定义了一整套复杂的标签,以描述调用的远程过程、参数、返回值和出错信息,等等。传统的基于 SOAP 的 Webservice 存在复杂性较高、效率较低的



问题。REST(REpresentational State Transfort)形式上表述为客户端通过申请资源来实现状态的转换,RESTful + JSON 轻量级的数据交换协议与 SOAP 和 XML 相比更加简洁,更适合于面向资源的应用。

2.3.1 RESTful SUSHI 接口、安全验证方式及过滤器

COUNTER R5 版的 SUSHI 接口文档托管于 SwaggerHub^[17],定义了 COUNTER_SUSHI 接口的技术规范,其中 4 个接口是明确要求实现的(见表 7)。

表 7 COUNTER_SUSHI 接口

接口	说明
GET /status	返回 COUNTER_SUSHI API 服务的当前状态。接口返回一条消息,其中包括 API 的运行状态、服务条目的 URL 以及服务备注信息
GET /reports	返回 COUNTER_SUSHI API 服务支持的报告列表。包括报告标识符、版本号、报告名称等信息
GET/reports/{ Report _ ID in lower case}	每个支持的报告都有自己的路径,例如使用 GET /reports/tr_b1 获取 TR_B1 报告(不包括 OA_Gold)
GET /members	返回联盟的成员或站点。返回值包括一系列客户账户信息,包括针对每个客户的 Requestor ID 和 Customer ID 等信息

COUNTER R5 规定了 3 种安全验证方式:①通过 Customer ID 和 Requestor ID 的组合验证;②通过 IP 地址验证;③通过调用接口时的 API key 验证。并且明确说明不可使用未在规范中定义的验证方式,这样简化了用户在调用 SUSHI 接口时的实现难度。

COUNTER R5 规定使用过滤器和属性对主报告进行定制需求时,需要将过滤器和属性明确包含在 COUNTER_SUSHI 请求中作为参数。以 Swaggerhub 的样例 SUSHI 服务为例,获取 2019 年 1 月份的 TR_J1 报告时,请求的 URL 格式如下: https://virtserver.swaggerhub.com/COUNTER/couter-sushi_5_0_api/1.0.0/reports/tr_j1?customer_id=szu&begin_date=2019-01-01&end_date=2019-01-31。如果接口状态正常,返回的报告数据包括报告头和报告内容(见图 1)。

```
"Report_Header": {
  "Created": "2019-09-08T22:47:31Z",
  "Created_By": "EBSCO Information Services",
  "Customer_ID": "12345",
  "Report_ID": "TR_J1",
  "Release": "5",
  "Report_Name": "Journal Requests (Excluding \"OA_Gold\")",
  "Institution_Name": "Mt. Laurel University",
  "Institution_ID": {
    {
      "Type": "ISNI",
      "Value": "1234 1234 1234 1234"
    }
  },
  "Report_Filters": [
    {
      "Name": "Begin_Date",
      "Value": "2019-01-01"
    }
  ],
  "Report_Attributes": [
    {
      "Name": "Attributes_To_Show",
      "Value": "Data_Type|Access_Method"
    }
  ],
  "Exceptions": [
    {
      "Code": 3040,
      "Severity": "Warning",
      "Message": "Partial Data Returned.",
      "Help_URL": "string",
      "Data": "Usage data has not been processed for all r"
    }
  ]
},
"Report_Items": [
  {
    "Title": "Journal of Economics",
    "Item_ID": {
      {
        "Type": "Print_ISSN",
        "Value": "0931-8665"
      }
    },
    "Platform": "EBSCOhost",
    "Publisher": "Springer",
    "Publisher_ID": {
      {
        "Type": "ISNI",
        "Value": "1234 1234 1234 1234"
      }
    },
    "Data_Type": "Journal",
    "Section_Type": "Article",
    "YOP": "2010",
    "Access_Type": "Controlled",
    "Access_Method": "Regular",
    "Performance": {
      "Period": {
        "Begin_Date": "2019-01-01",
        "End_Date": "2019-01-31"
      },
      "Instance": {
        "Metric_Type": "Total_Item_Requests",
        "Count": 21
      }
    }
  }
]
```

图 1 TR_J1 报告 JSON 格式内容

2.3.2 COUNTER R5 报告的交付方式

COUNTER R5 对于数据库商提供 SUSHI 服务的交付方式也提出了明确的要求:①数据库商需要在 Web 端的管理平台提供表格格式的报告,所有支持的报告必须可以通过 SUSHI 接口进行访问;②每份报告需要保存在独立的文件中;③必须每月提交报告;④数据必须在 4 周内提供;⑤每个月的统计数据必须完整,否则要使用异常代码 3031 来告知用户;⑥至少提供当前年份和之前 24 个月的统计报告,新加入的数据库商除外;⑦数据库商实现了新版的 COUNTER 规范,在提供新版 COUNTER 报告的同时,还要提供上一版本的 COUNTER 报告。

针对联盟的统计需求,COUNTER R5 明确规定联盟的管理员可以通过 Web 端管理后台获取其成员馆的统计报告,SUSHI 实现也必须支持/members 接口,通过该接口联盟管理员可以获取到联盟下所有成员馆的 SUSHI 访问账号。

2.4 COUNTER R5 与 COUNTER R4 相比的变化及兼容方式

COUNTER R5 更新了整个报告体系,重命名了计量类型,引入了新的元素来拓展规范的适用范围,引入了全新的 SUSHI 协议。两个版本常用的报告、计量类型之间的关系详见表 8。

相比于 COUNTER R4,COUNTER R5 的这些新变化使得统计报告内容上更加一致,获取途径上更加方便,图书馆可更准确地获取数字资源的使用数量,更准确地评估数字资源的使用成本,这对于图书馆的资源评价工作具有非常积极的意义。



表 8 COUNTER R5 常用报告和计量类型比较

	COUNTER R4	COUNTER R5	说明
报告	Book Report 1	Book Requests (ExcludingOA_Gold)	R5 中的计量类型 Unique_Title_Requests 与 BR1 中的 full-text 请求一致
	Journal Report 1	Journal Requests (ExcludingOA_Gold)	R5 的 Total_Item_Requests 代表用户的全部活动 Unique_Item_Requests 标识用户对资源的唯一访问数据
	Database Report 1	Database Usage	计量类型 Searches_Regular 保持不变
计量类型	ft_total	Total_Item_Requests	两个计量类型相同
	ft_html	Total_Item_Requests	R5 用 Total_Item_Requests 代替 R4 中特定格式的计量类型
	ft_pdf	Unique_Item_Requests	
	ft_ps	Unique_Title_Requests	
	search_reg	Searches_Regular Searches_Platform	数据库报告,使用 Searches_Regular。平台级别报告,使用 Searches_Platform

3 在 DRAA 使用统计模块中应用 COUNTER R5 的探索

DRAA 系统的使用统计模块于 2013 年投入使用,于 2014 年开始支持 COUNTER R4。DRAA 系统支持手工上传 COUNTER 报告和通过 SUSHI 方式自动获取 COUNTER 报告,从 COUNTER R4 开始明确要求数据商提供 SUSHI 服务,DRAA 系统从 COUNTER R4 开始只支持通过 SUSHI 服务的方式获取 COUNTER 报告。

3.1 DRAA 使用统计模块

DRAA 使用统计模块对于统计数据的管理主要分为 3 个层次:数据获取、数据处理、报告展示(见图 2)。DRAA 自主研发了基于 SOAP Webservice 的 SUSHI 客户端及 XML 文件解析器(见图 3),可以实现对 COUNTER R3 和 R4 SUSHI 服务器的调用,并处理了常见的安全认证需求,分析 COUNTER 报告数据结构,在数据库中建立了对应的数据模型,基于数据模型研发了 COUNTER 报告解析器。

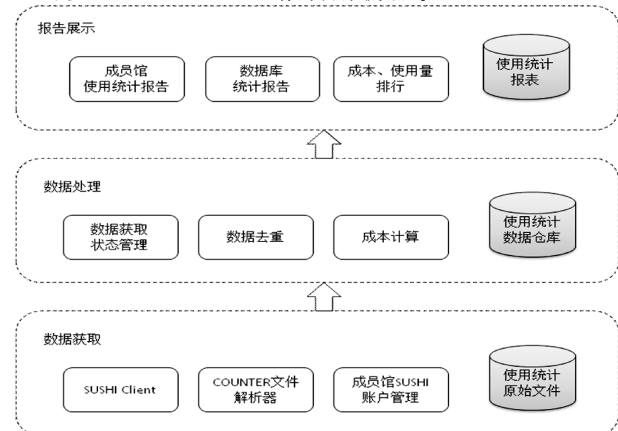


图 2 DRAA 统计模块系统结构

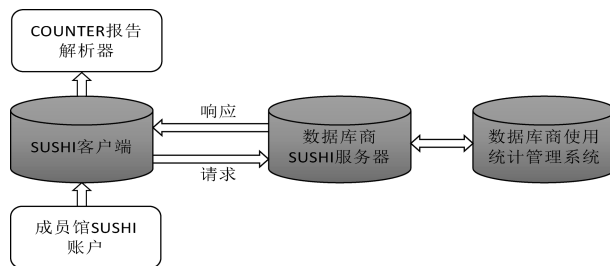


图 3 DRAA 统计模块自动收割 COUNTER 报告过程

在数据获取层面,系统统一配置了数据库商 SUSHI 服务器的参数,并向数据库商收集各个成员馆的 SUSHI 服务账号信息,定期获取成员馆的使用统计报告,同时系统会对所有成员馆的 SUSHI 账号定期检查维护。在获取到 COUNTER 报告之后,利用报告解析器将数据解析到对应的数据模型,并存储在数据库表中。通过 SUSHI 获取到的统计数据原始文件,也会保存在文件系统中,用户可下载这些文件供参考和研究。在数据处理层,系统对获取到的数据进行查询去重等操作,将统计数据与 DRAA 资源百科信息相关联,并结合数据库商上传的购买费用数据计算使用成本,依据成本及使用量等信息生成各类使用数据排行。最后将处理好的数据以 DRAA 单馆统计报告及数据库商报告的形式呈现给用户,在报告中给出数据库全部成员馆的平均下载成本、用户本馆的下载成本以及全文下载成本在全部订购成员馆中的排名,供成员馆比较本馆的电子资源使用情况。

DRAA 对引进数据库的经济效益评估,主要是针对检索成本、登录成本和全文下载成本^[18],在 DRAA 系统中用户比较关心的统计值就是检索量、访问量和全文下载量。在通过 SUSHI 获取数据时



主要是使用 DB1、JR1、BR1 这 3 种报告。

3.2 COUNTER Release 5 在 DRAA 的应用

DRAA 在应用 COUNTER R4 方面已经有了一些成功的经验^[19],为了在 DRAA 中应用 COUNTER R5,主要从两个方面进行了改进:一是让统计模块兼容 COUNTER R5 的数据模型,使 COUNTER R5 的统计数据可以融入现有的统计模块;二是新增 RESTful SUSHI 接口的客户端及文件解析器,让系统可以获取并解析到 COUNTER R5 的报告。

3.2.1 兼容 COUNTER R5 数据模型

COUNTER R5 新增的指标和属性中,主机类型、数据类型、章节类型、访问类型、访问方式等主要是对资源的描述,计量类型是对用户行为的描述。在 DRAA 系统中应该首先将 COUNTER R5 中的报告体系和计量类型映射到数据模型,根据 DRAA 引进数据库的经济效益评估需求,可以使用 DR_D1、TR_B1、TR_J1 这 3 个标准视图来获取数据库、电子书、电子期刊的使用统计数据,使用 Total_Item_Requests 和 Searches_Regular 计量类型来度量用户的访问行为。

对 COUNTER R5 的 JSON 格式数据模型进行分析,使用数据表以兼容 COUNTER R5 的数据模型,核心的数据表有 3 个,分别为描述资源的表、记录资源使用量的表以及记录计量类型的表。资源的描述表记录了资源名称、标识符、平台、出版社等与资源相关的信息,可以保存 COUNTER R5 JSON 格式报告中 Report_Item 节点的 Title、Platform、Publisher 等信息(见图 4)。Report_Item.Performance 节点记录了资源的具体的使用量,其中包括使用量的日期范围信息 Period,具体的使用量 Instance,其中 Instance 节点又包括计量类型 Metric_Type 和统计量的值 Count,可以使用资源使用量表来记录这些信息(见图 4)。计量类型的表使用行记录来保存所有的计量类型,在表中增加 COUNTER R5 中新增的计量类型如 Total_Item_Investigations 等,COUNTER R5 中与 R4 等价的计量类型可以继续使用现有的计量类型来保存数据,例如 Total_Item_Requests 类型的数据可以在系统中使用 ft_total 字段来记录(见图 5)。

Field Name	Field Type	Field Name	Field Type
infpid	uuid	pid	uuid
libid	uuid	infpid	uuid
orgid	uuid	performanceyear	integer
phystype	integer	performancemonth	integer
itemname	varchar(1000)	category	integer
itemplatform	varchar(1000)	metrictype	integer
itempublisher	varchar(1000)	requestcount	integer
itemdatatype	integer	createdtime	timestamp
itemidentifier	text	createdby	uuid
createdtime	timestamp	lastupdatedtime	timestamp
createdby	uuid	lastupdatedby	uuid
lastupdatedtime	timestamp		
lastupdatedby	uuid		

图 4 资源描述表与资源使用量表

typeid	typename	typedescription
1	ft_ps	Postscript file full text requests
2	ft_pdf	PDF file full text requests
3	ft_html	HTML full text requests
4	ft_total	Total full text requests
5	Toc	Table of Content Requests
6	abstract	Request for abstract view of article (detailed article metadata without full text)
7	reference	Number of views to bibliographic references pages associated with an article
8	data_set	Requests for supplementary data sets referenced in an article
9	audio	Requests for audio clips referenced in an article
10	video	Requests for video clips referenced in an article
11	image	Requests for images referenced in an article
12	podcast	Request for podcasts referenced in an article
13	search_reg	Regular searches as conducted by users (excludes searches attributed to a
14	search_fed	Searches conducted by federated search or automated search processes
15	session_reg	Count of normal sessions as conducted by users (excludes sessions attribut
16	session_fed	Count of sessions conducted by federated search or automated search proc
17	count	Simple count of usage (no breakdown) for the Category
18	other	The 'other' category was provided for within the original COUNTER schema

图 5 计量类型表

3.2.2 增加 RESTful SUSHI 接口的客户端

COUNTER R5 规定了数据库商必须实现 RESTful SUSHI 接口,并给出了接口的详细规范。在 Java 中实现对 RESTful API 接口的调用方式有很多种,例如使用 Jersey API、URLConnection、HttpClient、Spring RestTemplate 等。DRAA 系统计划使用 HttpURLConnection 方式来实现 COUNTER R5 的 RESTful SUSHI 客户端,核心代码如图 6 所示,URLConnection^[20]工具类继承自 URLConnection,包含在 JDK 的 java.net 包中,它提供了访问 HTTP 协议的基本功能,它对大部分工作进行了包装,屏蔽了不需要的细节,能够简化开发和后期维护。在获取到 JSON 格式的报数据之后,可以使用 JSONObject、Jackson ObjectMapper、GSON、FastJSON 等工具解析数据,获得的 JSON 格式的报告以.json 格式的文件存档。

```
URL realUrl = new URL(url);
conn = (URLConnection) realUrl.openConnection();
conn.setRequestMethod("GET");
conn.setConnectTimeout(20000);
conn.setReadTimeout(300000);
conn.setRequestProperty("Content-Type", "application/json");
is = conn.getInputStream();
br = new BufferedReader(new InputStreamReader(is));
```

图 6 调用 RESTful API



DRAA 系统的前端用户界面也需要兼容 COUNTER R5 的选项,在用户界面上应保持一致的风格。在系统后端根据用户设定的 SUSHI 服务的版本来选择对应的 SUSHI 客户端(见图 7)。在 COUNTER R4 和 R5 两个版本的过渡阶段,针对 COUNTER R4 的服务,使用 SOAP Webservice 获取数据,然后通过 XML 文件解析器解析;针对 COUNTER R5 的服务,使用 RESTful API 客户端获取数据,并使用 JSON 文件解析器解析;最后将来自 COUNTER R4 和 COUNTER R5 服务的数据汇总到数据库表中。

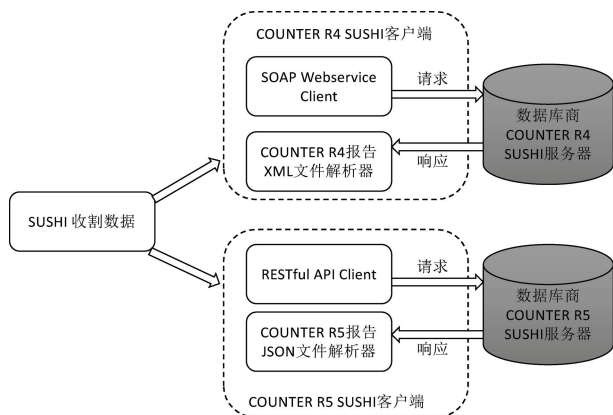


图 7 SUSHI 客户端结构图

4 结语

随着信息技术的发展,电子资源在载体形式和访问方式上都发生了巨大的变化,为了应对这些变化,COUNTER Release 5 引入了新的报告体系以及新的元素和属性,并且定义了全新的 RESTful SUSHI 接口,与 COUNTER Release 4 相比有了非常大的变化。目前各个数据库商还在努力实现 COUNTER Release 5 的合规性,提供符合 COUNTER Release 5 的报告必然是未来电子资源使用统计发展的方向。本文分析了 COUNTER Release 5 的新特征,对比分析了其相对于 COUNTER Release 4 的更新之处,为 DRAA 未来支持 COUNTER Release 5 做好理论上的准备和应用升级的计划,让 DRAA 对使用统计数据的应用紧贴时代发展趋势,走进技术前沿。

参考文献

- 1 李洪.新版 COUNTER 的特征及未来发展[J].中国图书馆学报,2012,38(6):29-37.
- 2 The COUNTER code of practice for release 4[EB/OL].[2019-02-20].<https://www.projectcounter.org/code-of-practice-sections/general-information/>.

- 3 The COUNTER code of practice for release 5[EB/OL].[2019-02-20].<https://www.projectcounter.org/code-of-practice-five-sections/abstract/>.
- 4 Organisations release 5 compliant and working toward compliance[EB/OL].[2019-02-20].<https://www.projectcounter.org/about/organisations-working-towards-release-5-compliance/>.
- 5 Oliver Pesch. COUNTER: looking ahead to release 5 of the COUNTER code of practice[J].The Serials Librarian,2016,71(2):83-90.
- 6 Oliver Pesch. COUNTER release 5: what's new and what it means to libraries[J].The Serials Librarian,2017,73:195-207.
- 7 胡大琴.中文数字资源使用统计数据的调查研究[J].图书馆情报工作,2017,61(15):104-110.
- 8 侯景丽.COUNTER R5 的新特性及对图书馆的影响[J].图书馆杂志,2018,37(12):46-55.
- 9 关于开通 SUSHI 自动收割 DRAA 成员馆使用统计服务的通知[EB/OL].[2019-02-24].<http://www.libconsortia.edu.cn/Acq-qothernotice/view.action?id=06d024ec-9c50-41ab-95a6-80356208fc59>.
- 10 Section 1.2: changes from COUNTER release 4, the COUNTER code of practice for release 5[EB/OL].[2019-02-24].<https://www.projectcounter.org/code-of-practice-five-sections/introduction-to-counter-code-of-practice-release-5/>.
- 11 Release 4 code of practice: usage reports[EB/OL].[2019-02-25].<https://www.projectcounter.org/codeofpractice-sections/usage-reports>.
- 12 Section 3.3: COUNTER report common attributes and elements, the COUNTER code of practice for release 5[EB/OL].[2019-02-25].<https://www.projectcounter.org/code-of-practice-five-sections/3-0-technical-specifications/#common-attributes>.
- 13 Simple Object Access Protocol (SOAP)[S/OL].[2019-03-01].<https://www.w3.org/TR/soap/>.
- 14 Extensible Markup Language (XML)[S/OL].[2019-03-01].<https://www.w3.org/TR/xml/>.
- 15 Representational state transfer[EB/OL].[2019-03-01].https://en.wikipedia.org/wiki/Representational_state_transfer.
- 16 JavaScript Object Notation[EB/OL].[2019-03-01].<https://www.json.org/>.
- 17 The specification for the RESTful COUNTER_SUSHI API[EB/OL].[2019-03-02].<https://app.swaggerhub.com/apis/COUNTER>.
- 18 肖珑,章琳.引进数据库的发展趋势与价格成本策略[J].大学图书馆学报,2015,33(1):5-13+20.
- 19 Lan Ye, Wei Yang, Weiming Lin. DRAA e-resources usage statistics services in China: research and practice[J].The Electronic Library,2018,36(6):1043-1061.
- 20 Class HttpURLConnection[EB/OL].[2019-03-04].<https://docs.oracle.com/javase/8/docs/api/java/net/HttpURLConnection.html>.

作者单位:深圳大学图书馆,广东深圳,518086

收稿日期:2019年5月30日

(责任编辑:关志英)

(转第 41 页)