



Usage Analysis of University of Manitoba Licensed Databases via EZproxy's Logs

Wei Xuan

Abstract:

Compared to vendor-provided data, Transaction Log Analysis (TLA) can provide unique benefits to a library in analyzing database usage. Meanwhile, it also requires librarians to have a broad knowledge of Information Technology in order to implement a TLA. This article will present an in-house database system developed at the University of Manitoba Libraries to analyze EZProxy's Starting Point Uniform Resource Locator logs. The database visits that happened in the past two years have been imported into this system and are split into two categories, on-campus and off-campus visits. Unique statistical information from each category is also discussed.

To cite this article:

Xuan, W. (2019). Usage analysis of University of Manitoba licensed databases via EZproxy's logs. *International Journal of Librarianship*, 4(2), 76-93.

To submit your article to this journal:

Go to <http://ojs.calaijol.org/index.php/ijol/about/submissions>

Usage Analysis of University of Manitoba Licensed Databases via EZproxy's Logs

Wei Xuan

University of Manitoba Libraries, Winnipeg, Manitoba, Canada

ABSTRACT

Compared to vendor-provided data, Transaction Log Analysis (TLA) can provide unique benefits to a library in analyzing database usage. Meanwhile, it also requires librarians to have a broad knowledge of Information Technology in order to implement a TLA. This article will present an in-house database system developed at the University of Manitoba Libraries to analyze EZProxy's Starting Point Uniform Resource Locator logs. The database visits that happened in the past two years have been imported into this system and are split into two categories, on-campus and off-campus visits. Unique statistical information from each category is also discussed.

Keywords: Database Usage Analysis, Transaction Log Analysis (TLA), EZProxy, SPULog

INTRODUCTION

A database is “a structured set of data held in computer storage and typically accessed or manipulated by means of specialized software”(OED, 2019). To academic libraries, the term database represents searchable online resources. A database may provide users access to abstracts, or full-text articles, or images, etc. The access is usually paid by the library. Since the start of migrating content from print materials to online databases, libraries have been looking for methods to better analyze the usage of licensed e-resources. The analysis of database usages should allow librarians to know how different databases are used by different groups of users and what the user groups preferred. Discovery tool and information sources are most likely chosen when users perform information searches. With the knowledge of such information-seeking behaviors, librarians will be able to provide custom services to particular user groups. On the other hand, analyzing database usages also allows libraries to better estimate the value of a particular database to their institutions, such as whether or not it is a good investment to subscribe to a database. Especially with more universities moving to a faculty-oriented budget model and when budget cuts happen more frequently, the need to demonstrate how licensed databases are used by faculties is becoming more urgent in the decision-making process.

This article analyzed the usage of licensed databases at University of Manitoba for a period of two years (September 1st, 2016 – August 31st, 2018). A typical EZProxy Uniform Resource Locator (URL) is like <https://abclib.idm.oclc.org/login?url=https://www.jstor.com> . The first part of the URL is the proxy prefix (<https://abclib.idm.oclc.org/login?url=>), which is the EZProxy

server's URL, and the rest is the target URL (<https://www.jstor.com>), which is the unique database URL. The proxy prefix and the target URL make a starting point URL (SPU) (OCLC, 2018b). The proxy prefix stays the same for all resources licensed by an institution. The target URL represents the specific e-resource that a user visited. This database usage analysis is based on the EZProxy SPU logs as the target URLs are recorded each time a starting point URL is clicked (OCLC, 2019). Compared to normal EZProxy logs, SPU logs greatly simplify the analysis process and allow librarians to focus on analyzing how the user accesses a particular database URL without being distracted by information, such as text, images, etc., transferred between the database website and the EZproxy server.

The author designed a MySQL database and imported EZProxy's SPU logs into it. This database also contains the University of Manitoba (UofM) campus Internet Protocol (IP) ranges, UofM Identity (ID) for faculty, staff and students, and all database records from the Libraries' AZ Database list on LibGuide. With this database, the author was able to identify the following:

1. Where was a database visit made? Was it from campus or off-campus?
2. How a database visit was initiated? Did the user start from the Libraries' AZ Database List or the discovery platform - Primo?
3. How frequently was a database visited during a specific period? Regarding on-campus visits, how many visits were from the wireless network and how many from Ethernet in a particular building, such as the library building? Regarding off-campus visits, how many visits were made from a particular user group (faculty, graduate students, and undergraduate students)?
4. What was the preference for a particular user group when they visited databases off-campus? For example, what are the top 10 databases used by graduate students when they are off-campus?

Thus, the database analysis method described in this article allows the author to investigate both the usage of a particular database in the period and the off-campus information-seeking activities of a particular user group in the time frame.

LITERATURE REVIEW

There are generally two major methods for librarians to know how their subscribed databases are used among users. One is through vendor-supplied usage data and the other is through transaction log analysis (TLA)(Peters, 1993). Each method has its advantages and disadvantages (Duy & Vaughan, 2003).

With the vendor-supplied usage data, the issues can be generalized as a lack of credible and comparable data. During the early implementation, different vendors were using various types of measures in their reports. Due to the lack of context for the data, such as how the number was counted and collected, librarians had great difficulty in comparing and interpreting the usage data (Blecic, Fiscella, & Wiberley, 2001). To solve the problem, an international organization, COUNTER (Counting Online Usage of Networked Electronic Resources), was formed to promote a consistent way for e-resource providers to generate usage reports (COUNTER, 2019). By releasing the Code of Practice, COUNTER is promoting standards in this area. The consistency of

usage data from the vendors that are COUNTER-compliant has been improved. However, even COUNTER-compliant vendors still might be able to count searches differently without violating the Code of Practice (Blecic, Fiscella, & Wiberley, 2007). Besides, since it is not mandatory, some vendors haven't applied the COUNTER Code of Practice.

The challenge for librarians with the COUNTER reports is that it is very difficult and time-consuming to put together reports from all vendors. To a research library that usually subscribes to hundreds of databases, it becomes an impossible mission for their librarians to group all the COUNTER data by themselves. To provide a solution, the National Information Standards Organization (NISO) initiated a working group to work on Standardized Usage Statistics Harvesting Initiative (SUSHI) in 2005 and approved the SUSHI standard in 2007 (NISO, 2019). It automates the transfer of usage data from vendors to a central system, whether it is the Integrated Library System (ILS) or a reporting portal.

Even with COUNTER and SUSHI, there are still disadvantages to solely rely on vendor-provided usage data. The biggest issue is that the vendor cannot have the user's information. The most commonly used authentication method for a database visit request is still the IP authentication. Approving a request simply based on whether it's from a registered IP address, the database provider is unable to include user information in the statistics. As a result, a database vendor can provide the number of searches but cannot answer questions like who made those searches. Even with the Security Assertion Markup Language (SAML) authentication, such as Shibboleth (Shibboleth, 2019), the user's attribute released to the database provider may only contain basic group information, such as faculty and student. Secondly, as library consortium becomes the trend in managing resources and negotiating licenses (Liu & Fu, 2018), member libraries in a consortium may lose statistical information about their usage. This issue has been existing in the public library environment for a long time. Thirdly, it is mentioned by Wan and Liu (Wan & Liu, 2010) that a library may have more than one account with a database provider.

Despite the inconveniences in utilizing vendor-supplied usage data, a survey carried out in 2006 (Baker & Read, 2008) found that the majority of research libraries relied on it to make subscription decisions, justify expenditures and meet reporting requirements.

TLA is to analyze the transaction logs of a proxy server that users have to use to access the licensed e-resources. If an e-resource visit can bypass the proxy server, such as campus IP addresses directly recognized by the database provider or SAML authentication on the database provider's website, that visit will not be recorded in the proxy server's log. Thus, the library will lose certain statistical information about those kinds of visits when performing TLA. The advantage of TLA is that librarians have first-hand access to raw data. If well analyzed, those logs will provide a better demonstration of user activities. However, the difficulty with TLA is how to retrieve useful information from the log, as a web server log usually contains all kinds of web activities and is saved as a text file, and how to group log files, as log files are usually created daily. Those challenges require librarians to either have a strong background in Information Technology (IT) or work closely with IT staff in order to understand and analyze transaction logs.

Coombs conducted a project at the State University of New York College at Cortland (SUNY Cortland), which analyzed EZProxy transaction logs (Coombs, 2005). EZProxy is the

commercial product that SUNY Cortland used to access databases. This project demonstrated that librarians could use TLA to study user information behavior on accessing the library's e-resources. Unfortunately, the author didn't demonstrate whether their method was able to track how the database visits change with time. The author also didn't give many details about how their system used to do TLA was designed and developed and how EZProxy log files were imported into this system. Without such information, it would be very hard for other institutions to reproduce the work.

Gonzales published a way to automatically parse EZProxy's SPU log (Gonzales, 2018). Since OCLC doesn't provide any tools to analyze EZProxy transaction logs and the commercial tools on the market are not optimized for EZProxy, Gonzales developed a Python script to import raw data from EZProxy SPU logs into a single CSV file and used two additional Python modules, *Pandas* and *Matplotlib*, to analyze it. The issue with this method is that the analysis of data stored in a CSV file is limited by the function of these two Python modules. Considering the number of records stored in the CSV file, it would cause a great load for a workstation to open it, not to mention to run any possible queries to it.

In this article, the author presented the database developed at the University of Manitoba Libraries to analyze EZProxy SPU logs. EZProxy SPU logs generated after September 1st, 2016, have been being imported. The author analyzed all the visits to licensed e-resources within a period of two years (Sept. 1st 2016 – August 31st, 2018). In this article, "database platform" means the e-resource provider's website where users can search articles, images, etc. Several databases may share one database platform. For example, EBSCO has many databases, such as *Academic Search Complete*, *Ageline*, *Alternative Press Index*, etc. All the databases are using the same database platform, *search.ebscohost.com*. *Alexander Street Press* also hosts several databases. However, the websites for the databases are different. For example, the website for the *American Film Scripts Online* database is <http://solomon.afso.alexanderstreet.com/> and the website for the *Asian American Drama* database is <http://solomon.aadr.alexanderstreet.com>. They will be treated as different database platforms in the analysis.

If a user goes to the Libraries' AZ database list on LibGuide and clicks the link to visit a specific database, in this article, the visit is treated as from the AZ database list. If a user uses the library's discovery platform, *Primo* from *Ex Libris*, to find an article and clicks the link to access it on the database platform, the visit is treated as from Primo.

METHODOLOGY

At the University of Manitoba Libraries, the licensed databases are set up to only allow visits from the EZProxy server. Thus, the EZProxy log records all the starting point URLs in the visits to our licensed e-resources, no matter whether the visit is from campus or off-campus.

When receiving a database-visit request, EZProxy will first check the requester's IP address. UofM IP ranges are whitelisted by the EXProxy server. If the request is from an UofM IP address, which means the user is on campus, EZProxy will automatically grant access to the user. In the log file, the word "auto" is recorded as the user identity (ID) for this visit. If the request is not from a UofM IP address, which means the user is off-campus, EZProxy will direct the user to a login

page where she/he needs to use the library ID to authenticate. For UofM faculty, staff and students, EZProxy uses SAML 2.0 protocol to authenticate them against signUM which is a central identity system used at UofM. In EZProxy log, the user’s UMnetID, which is the user’s unique identifier in signUM, will be recorded as the username for this visit. In both cases, the user’s IP address will be stored in the EZProxy log file.

We are using the following value for the EZProxy LogSPU(OCLC, 2018a) directive.

```
LogSPU -strftime /log/preprod/ezlogs/spu%Y%m%d.log %Y/%m/%d:%H:%M:%S)t\t%h\t%u\t%(ezproxy-spuaccess)i\t%v\t%U
-strftime is to record starting point URL information in a new log file each day
%t is the Date/time of request;
```

```
%h is the IP address of the host accessing EZProxy
%u is Username used to log into Ezproxy
%v is the hostname of the webserver
%U is the requested URL
```

A typical record in EZProxy log is like

```
2016/09/07:15:54:01      140.193.167.125  auto    proxy    www.jstor.org
http://www.jstor.org/action/showPublication?sid=primo&journalCode=science
```

A script has been developed to automatically import EZProxy SPU logs into a table, called “EZProxySPULog”, in a MySQL database, which also contains information about user groups, IP addresses on campus and A-Z database list that is used in our subject guide. The Entity-Relationship Diagram is presented in Figure 1.

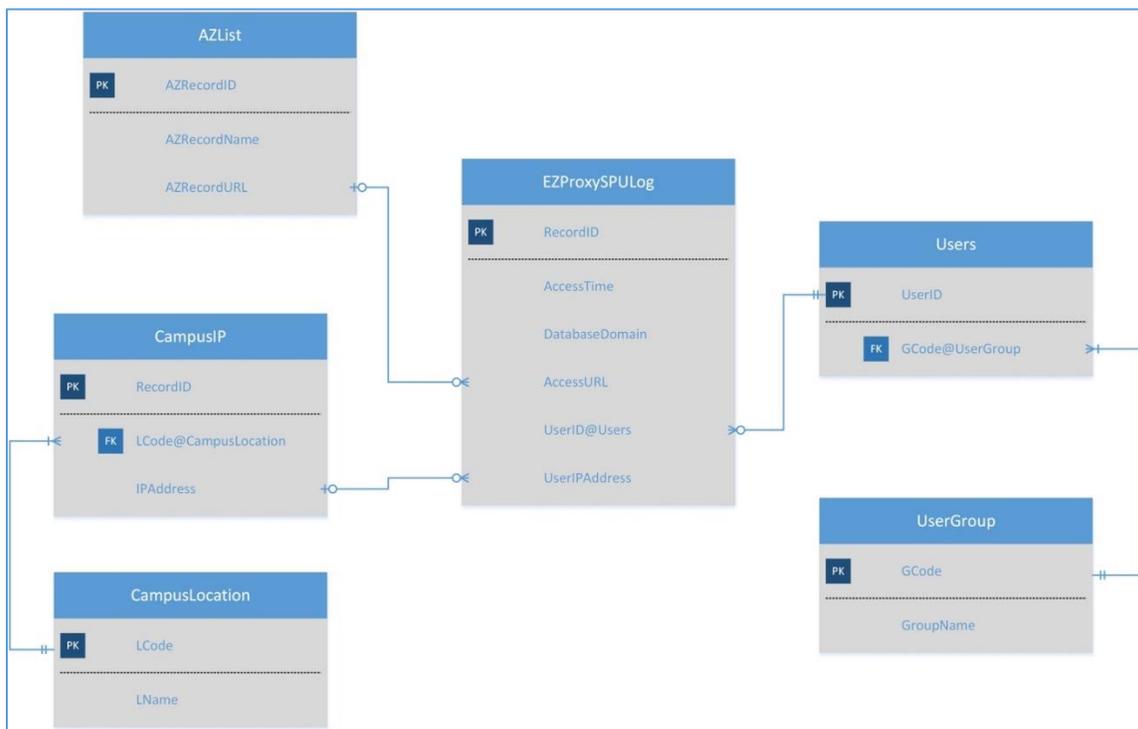


Figure 1. Entity Relationship Diagram for the database

The EZProxyLog table splits each visit record in the EZProxy log into five different fields, *Access Time, Database Domain, Access URL, User ID, and User IP Address*. The above log record is stored in the EZProxySPULog table in the following way (Table 1).

Table 1. Sample of a record in the EZProxySPULog table

id	datetime	host	URL	user_id	IP
201609073592	2016-09-07 15:54	www.jstor.org	http://www.jstor.org/action/showPublication?sid=primo&journalCode=science	auto	140.193.167.125

The SQL query to retrieve the amount of on-campus/off-campus/total database accesses is

```

/*total accesses*/
SELECT COUNT(RecordID) AS TotalVisits
FROM EZProxySPULog
WHERE datetime BETWEEN @StartTime AND @EndTime;

/* on-campus accesses*/
SELECT COUNT(RecordID) AS OnCampusVisits
FROM EZProxySPULog
WHERE user_id='auto'
AND datetime BETWEEN @StartTime AND @EndTime;

/* off-campus accesses*/
SELECT COUNT(RecordID) AS OffCampusVisits
FROM EZProxySPULog
WHERE user_id!='auto'
AND datetime BETWEEN @StartTime AND @EndTime;

```

The Libraries Systems department at the University of Manitoba Libraries gets a list of all UofM students from the IT department each semester. The list contains student name, UMnetID, the affiliated faculty and contact information. Graduate students are affiliated with Faculty of Graduate Studies on the list. Undergraduate students are affiliated with their respective faculty, such as Faculty of Science, Faculty of Engineering, etc. Data from this list is desensitized and only UMnetID and the affiliated faculty are used in the following two tables, *Users* and *Usergroup*, in the database. Since each student's UMnetID is unique, it is used for the UserID field, which is the primary key in the table "Users". There are 29 major academic units (faculty, college, school, etc.) that a student may be associated with at the university. These 29 units are stored in the UserGroup table. Each unit is a group in this database and assigned a 3-digit number as an identifier. Regarding other users, such as faculty, staff, alumni, retirees, etc., since they are in respective user groups in our Integrated Library System (ILS), we use their group information from ILS in this database. Groups 1-15 and 17-29 are used for undergraduate students. Group 16 is for graduate students. All UofM faculty and staff, including visiting scholars, are in a different group (GCode:100). Unfortunately, we cannot know the rank (Professor, Associate Professor, Assistant Professor, etc.) of a specific user in Group 100 as such information is not used in the ILS. We also don't know which faculty a specific user in Group 100 is affiliated with. The UserGroup table is attached in Appendix I. Alumni have access to a small portion of our licensed databases. Local community users have no remote access to our e-resources. Thus, no specific user groups were created in the UserGroup table for these two user categories. Since UserID is recorded in the SPU log for off-campus accesses (Table 2), we can combine three tables, *EZProxySPULog*, *Users* and *UserGroup*,

in the database, to analyze which user group made what kind of database access.-Some SQL queries are presented in the following.

Table 2. Sample of an off-campus visit in the EZProxySPULog table

id	datetime	host	URL	user_id	IP
2016082022	2016-08-20 0:18	www.jamaevidence.com	http://www.jamaevidence.com/resource/523	A*****	2*.*.*

Note: user_id and IP are masked by the author.

```
/*number of accesses filtered by group*/
SELECT UserGroup.GName AS UserGroup, COUNT (EZProxySPULog.RecordID) AS NumberOfVisits
FROM   EZProxySPULog INNER JOIN Users
ON     EZProxySPULog.UserID@Users= Users.UserID
INNER JOIN   UserGroup
ON     Users.GCode@UserGroup= UserGroup.GCode
WHERE  datetime BETWEEN @StartTime AND @EndTime
GROUP BY UserGroup
ORDER BY NumberOfVisits DESC;
```

A database A-Z list is presented to users on the Libraries' website. Database names and domains from this list are also inputted into the AZList table in this database. If a record in the AccessURL field in the EZProxyLog table matches a record in the AZRecordURL field in the AZList table, this visit is regarded as being initiated from the A-Z list on the Libraries' website as the user has to either click the database link on the A-Z list or use a bookmarked URL which is originally from the A-Z list.

UML is using Ex Libris' product, *Primo*, as the discovery platform. The article's hyperlink on Primo will direct EZProxy to proxy the article's URL from the database platform. A typical request EZProxy receives from Primo is like the following URL:

<https://uml.idm.oclc.org/login?&url=http://www.jstor.org/openurl?sid=primo&volume=335&aulast=Ponomarenko&date=2012&spage=947&issn=00368075&issue=6071>

In this case, the AccessURL field for this access will be the article's full URL

<http://www.jstor.org/openurl?sid=primo&volume=335&aulast=Ponomarenko&date=2012&spage=947&issn=00368075&issue=6071> instead of the database's URL <http://www.jstor.org>. Thus,

if a value in the AccessURL field in the EZProxySPULog table matches a value in the AZRecord field in the AZList table, the access is from the AZ database list. Otherwise, access is from the library's discovery platform.

```
/*-----Top 10 databases based on visit from AZ list----*/
SELECT AZList.AZRecordName AS DatabaseName, COUNT( EZProxySPULog.RecordID) AS NumberOfVisits
FROM   EZProxySPULog LEFT JOIN AZList
ON     EZProxySPULog.AccessURL= AZList.AZRecordURL
WHERE  EZProxySPULog.AccessTime BETWEEN @starttime AND @endtime
AND    AZList.AZRecordName IS NOT NULL
GROUP BY DatabaseName
ORDER BY NumberOfVisits DESC
LIMIT 10;
```

```
/*-----Top 10 database platforms based on visit from Primo----*/
SELECT EZProxySPULog.host AS DatabasePlatform, COUNT( EZProxySPULog.id) AS NumberOfVisits
FROM   EZProxySPULog LEFT JOIN AZList
```

```

ON      EZProxySPULog.AccessURL= AZList.AZRecordURL
WHERE  EZProxySPULog.AccessTime BETWEEN @starttime AND @endtime
AND    AZList.AZRecordName IS NULL
GROUP BY DatabasePlatform
ORDER BY NumberOfVisits DESC
LIMIT 10;

```

The University of Manitoba has its IP ranges. The IP ranges can be split into two major categories, Wi-fi and Ethernet. Regarding the Ethernet category, it can be further broken down into each building, such as *the Library building, the Education building, etc.* Regarding the Wi-fi category, a special range of IP addresses is reserved for graduate students, undergraduate students, and guest users. Considering the number of guest users is very small, most of the database visits from this special IP range are made from students. Thus, this IP-location information is stored in two tables in this database, *CampusIP*, and *CampusLocation*. For a record in the *EZProxyLog* table, if the value of the *UserID@Users* field is "auto", which means the visit is from a UofM IP address, we can check the location information. For example, if the IP address belongs to the Ethernet category, is it from the library building, the Science building or somewhere else? If the IP address belongs to the Wi-fi category, is it from the student and guest range? A sample query is presented in the following to group on-campus accesses based on location.

```

/*-----on-campus access grouped by location-----*/
SELECT CampusLocation.LName AS LocationName, COUNT(EZProxySPULog.RecordID) AS NumberOfVisits
FROM EZProxySPULog LEFT JOIN CampusIP
ON      EZProxySPULog.UserIPAddress = CampusIP.IPAddress
INNER JOIN      CampusLocation
ON      CampusIP.LCode@CampusLocation=CampusLocation.LCode
WHERE  CampusLocation.LName IS NOT NULL
AND    EZProxySPULog.AccessTime BETWEEN @starttime AND @endtime
GROUP BY      LocationName
ORDER BY      NumberOfVisits DESC
LIMIT 10;

```

When combining all the tables in the database together, we will be able to study complicated usages such as:

1. How often is a specific database on the AZ list accessed by users located in a particular building on campus?

```

SELECT AZList.AZRecordName AS DatabaseName, CampusLocation.LName AS LocationName
COUNT( EZProxySPULog.RecordID) AS NumberOfVisits
FROM  AZList RIGHT JOIN EZProxySPULog
ON    AZList.AZRecordURL = EZProxySPULog.AccessURL
LEFT JOIN CampusIP
ON    EZProxySPULog.UserIPAddress = CampusIP.IPAddress
INNER JOIN      CampusLocation
ON    CampusIP.LCode@CampusLocation=CampusLocation.LCode
WHERE  AZList.AZRecordName = @DatabaseName
AND    CampusLocation = @LocationCode
AND    EZProxySPULog.AccessTime BETWEEN @starttime AND @endtime

```

2. For users in a specific group, what database do they most frequently visit through the library's discovery platform when they are off campus?

```

SELECT EZProxySPULog.host AS DatabasePlatform, COUNT( EZProxySPULog.id) AS NumberOfVisits
FROM  AZList RIGHT JOIN EZProxySPULog
ON    AZList.AZRecordURL = EZProxySPULog.AccessURL
INNER JOIN Users

```

```

ON     EZProxySPULog.UserID@Users= Users.UserID
INNER JOIN  UserGroup
ON     Users.GCode@UserGroup= UserGroup.GCode
WHERE  EZProxySPULog.AccessTime BETWEEN @starttime AND @endtime
AND    AZList.AZRecordName IS NULL
AND    UserGroup.GCode=@UGroupCode
GROUP BY DatabasePlatform
ORDER BY NumberOfVisits DESC
LIMIT 5;
    
```

RESULTS

During the period analyzed in this report (September 1st, 2016 – August 31st, 2018), there were 5,061,398 visits to our licensed databases. Categorized by user location, 1,865,251 visits were made on-campus while the rest are 3,196,147 visits from off-campus (Figure 2). How the on-campus and off-campus visits change with time can be found in Figure 3.

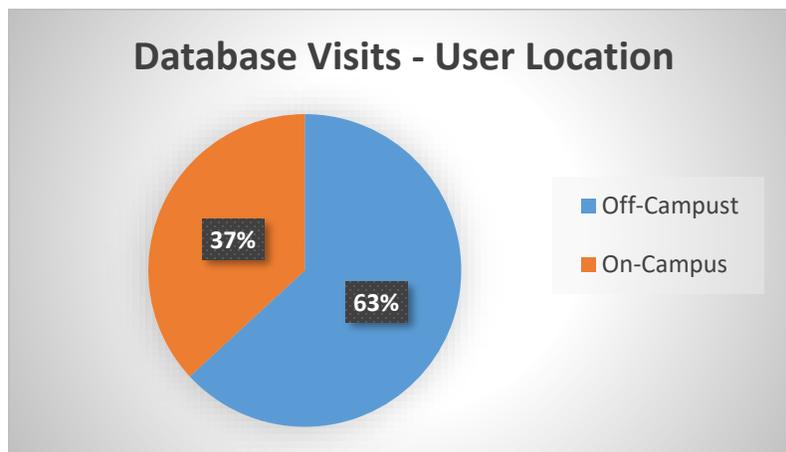


Figure 2. Database visits based on user location

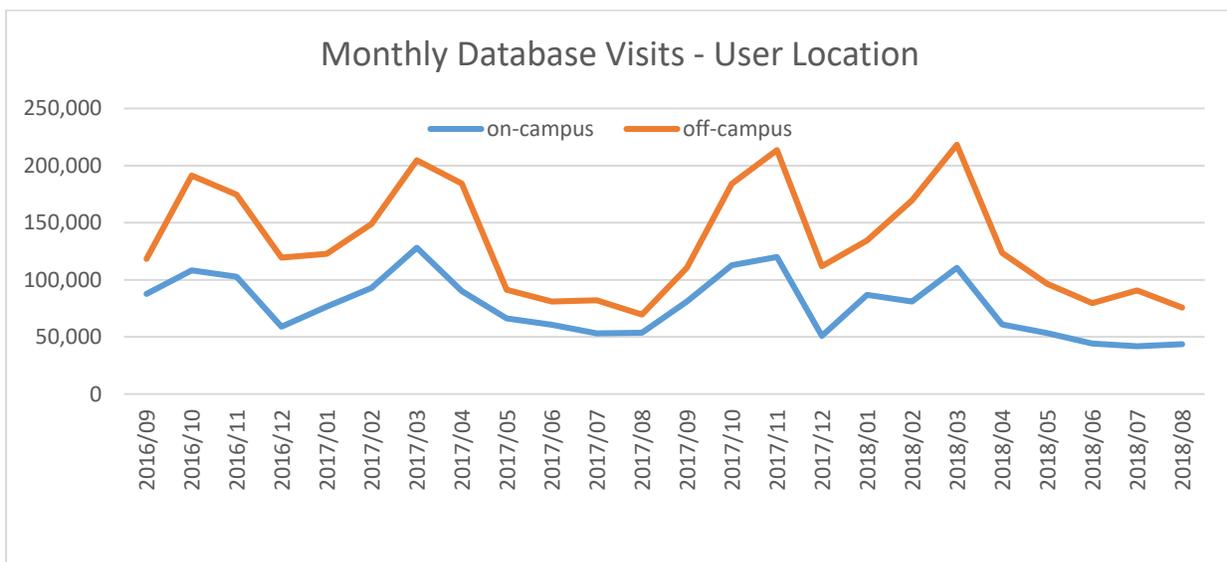


Figure 3. Monthly database visits – user location

Categorized by discovery channel, 844,090 visits were made from the AZ database list; 4,217,308 visits were made from Primo (Figure 4). How many times each of the two channels were used in a specific month can be found in Figure 5.

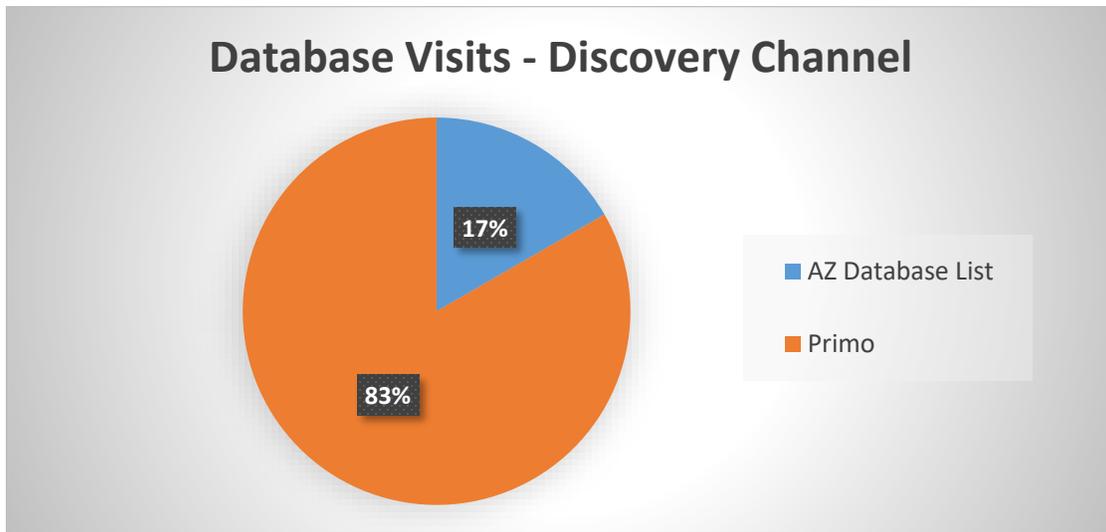


Figure 4. Database visit based on the discovery channel

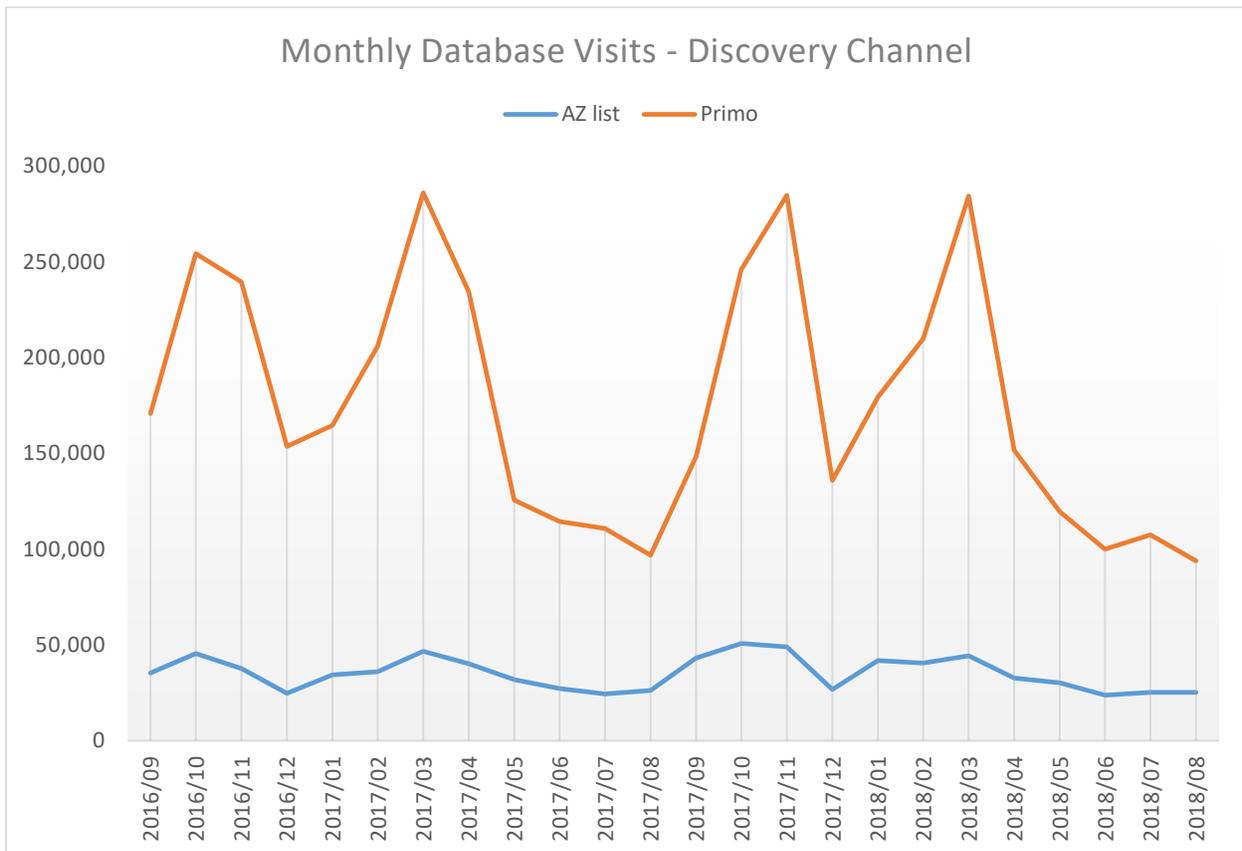


Figure 5. Monthly database visits - discovery channel

Regarding the on-campus visits, 237,854 were made from the AZ database list while 1,627,397 were made from Primo (Figure 6). The monthly visits through these two channels can be found in Figure 7.

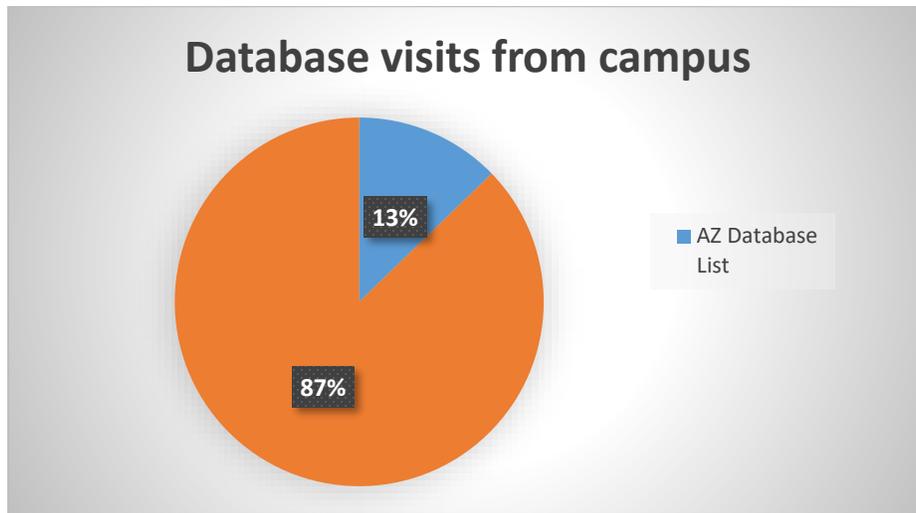


Figure 6. Database visits from campus

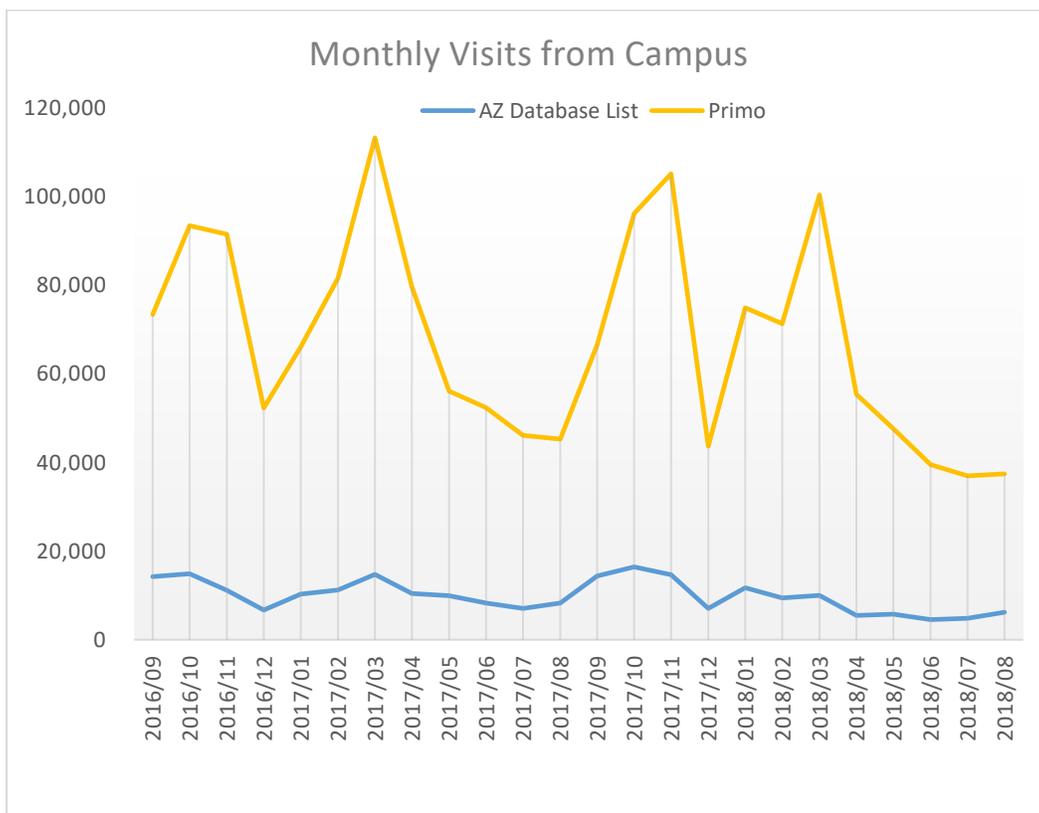


Figure 7. Monthly visits from campus

Regarding the on-campus visits from the AZ database list, 116,972 visits were made through the University of Manitoba Campus Ethernet. 72,853 visits were made through the WiFi network used by students and guests. The rest 48,029 visits were made through other methods such as some remote libraries on the UM network (Figure 8).

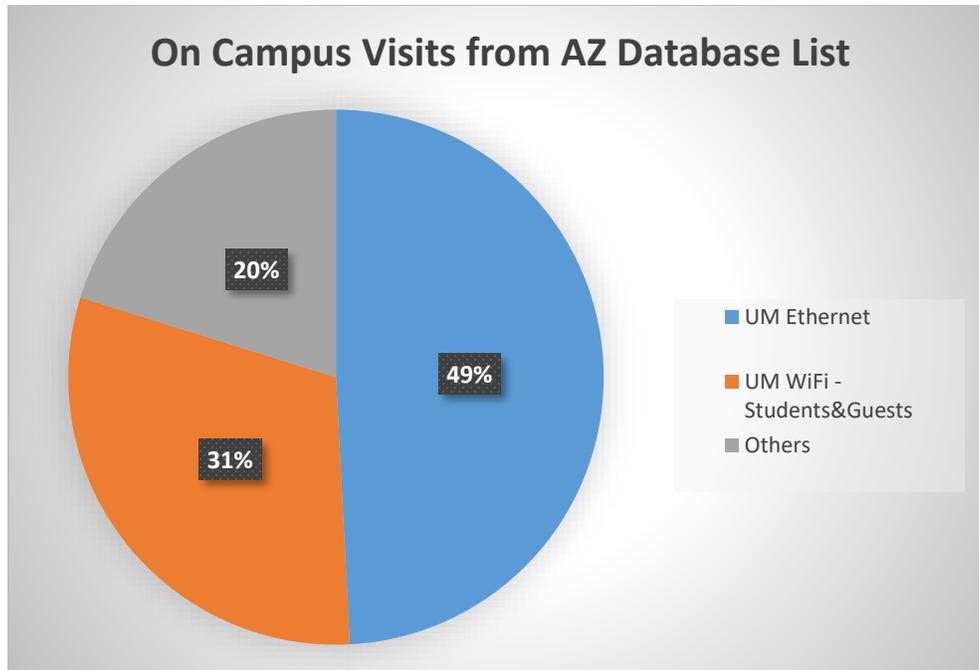


Figure 8. On-campus visits from the AS database list

Regarding the on-campus visits from Primo, 815,864 visits were made through the University of Manitoba Campus Ethernet. 659,817 visits were made through the WiFi network used by students and guests. The rest 151,716 visits were made through other methods such as some remote libraries on the UM network (Figure 9).

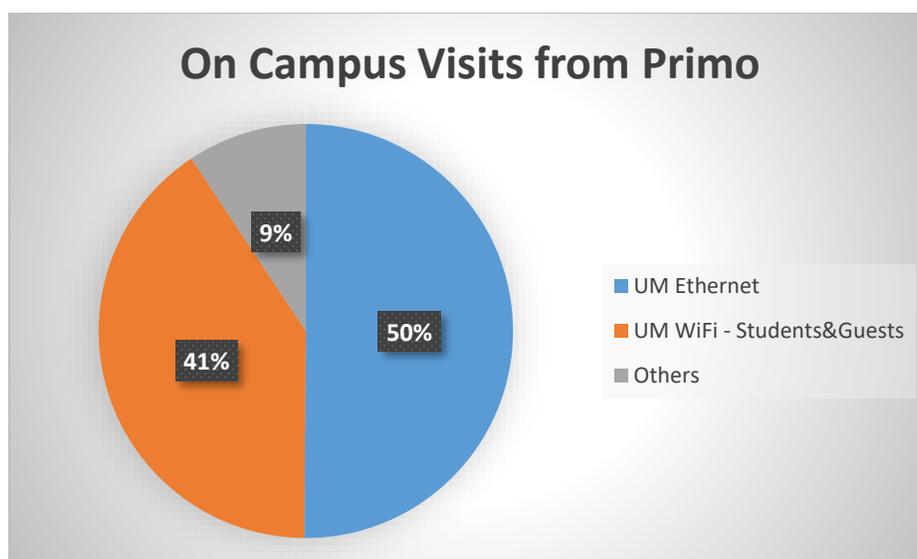


Figure 9. On-campus visits from Primo

The overall top 10 database domains that were most frequently visited from campus can be found in Table 3. Although dx.doi.org is not a database provider, the importance of digital object identifier (DOI) can be demonstrated by the number of visits to articles with a DOI.

Table 3. Top 10 database domains most frequently visited from campus

Database Domain	Total Visits
www.ncbi.nlm.nih.gov	116,764
www.sciencedirect.com	110,951
onlinelibrary.wiley.com	93,213
search.proquest.com	90,008
search.ebscohost.com	87,825
dx.doi.org	79,923
link.springer.com	65,473
scholar.google.com	62,192
www.tandfonline.com	61,951
openurl.ebscohost.com	49,189

Regarding the off-campus visits, 606,236 were made from the AZ database list while 2,589,911 were made from Primo (Figure 10. Database visits from off-campus

). The monthly visits from off-campus can be found in Figure 11. Although Primo is still the user’s first choice, compared to on-campus usage, more users choose AZ Database List to start their database access.

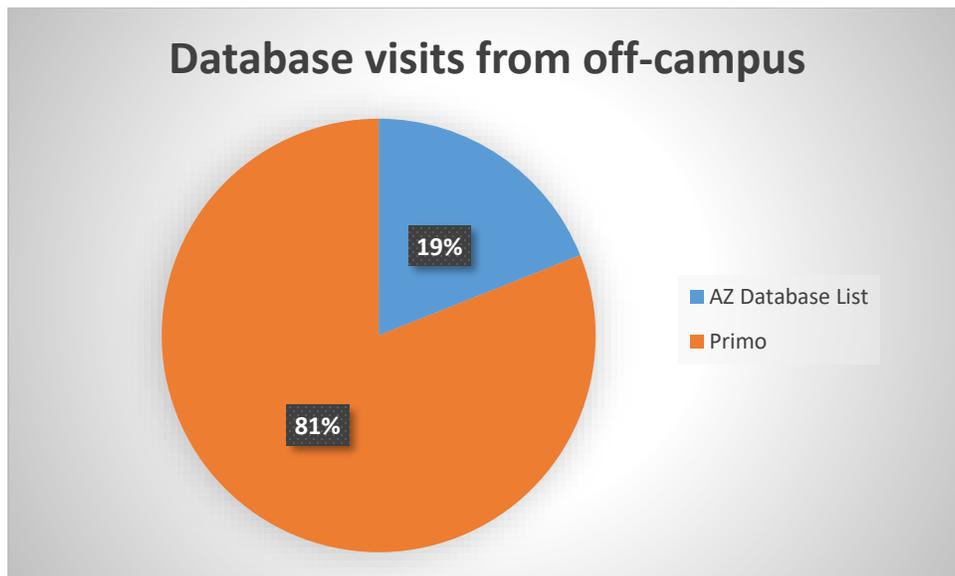


Figure 10. Database visits from off-campus

The top 10 user groups that made the most database visits from off-campus can be found in Table 4.

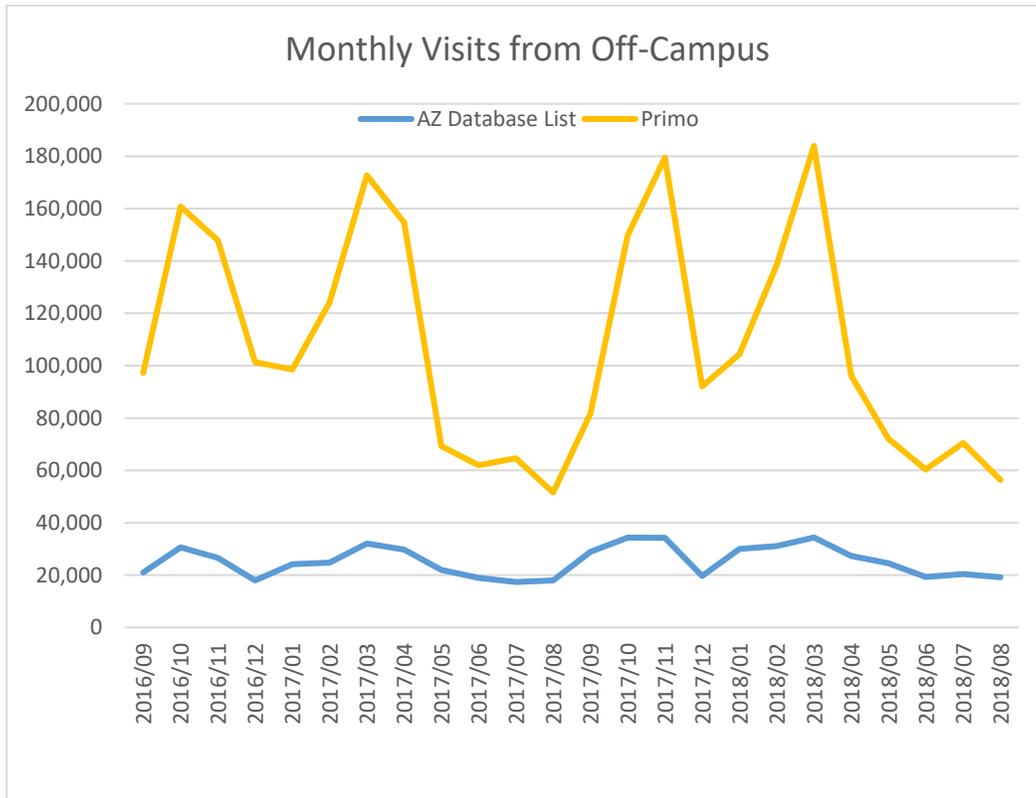


Figure 11. Monthly visits from off-campus

Table 4. Top 10 user groups from off-campus

User Group	Number Of Visits
Graduate Students (GCode :16, Faculty of Graduate Studies)	709,249
UofM Faculty and Staff (GCode: 100)	429,197
Undergraduate Students from Faculty of Arts (GCode: 1)	301,185
Undergraduate Students from Faculty of Science (GCode: 2)	209,722
Undergraduate Students from College of Medicine – PGME (GCode: 25)	204,032
Undergraduate Students from College of Medicine (GCode: 5)	113,398
Undergraduate Students from College of Pharmacy (GCode: 11)	107,163
Undergraduate Students from College of Nursing (GCode: 13)	105,023
Undergraduate Students from Faculty of Social Work (GCode: 12)	94,565

Undergraduate Students from Faculty of Education (GCode: 8)	69,215
---	--------

If we combine all undergraduate students together, they made the most database visits from off-campus. In this study, undergraduate students are affiliated with their respective faculty. Thus, in the current environment, graduate students made the most off-campus visits to the databases within the past two years. The author was interested in what databases were mostly used by these users. The top 10 databases on the AZ database list that were most frequently used by graduate students can be found in Table 5. The top 10 database domains that graduate students visited from Primo can be found in Table 6.

Table 5. Top 10 databases on the AZ database list visited by graduate students from off-campus

Database Name	Total Visits
PubMed	24,108
UpToDate	15,669
CINAHL with Full Text	6,378
Scopus	5,444
PsycINFO	4,285
Web of Science (all databases)	3,160
RxTx (CPA)	3,078
MEDLINE (Ovid)	2,453
LexiComp Online	2,310
ClinicalKey	1,677

Table 6. Top 10 database domains visited from Primo by graduate students from off-campus.

Database Domain	Total Visits
www.sciencedirect.com	48,975
onlinelibrary.wiley.com	45,381
dx.doi.org	44,062
www.tandfonline.com	37,676
scholar.google.com	36,239
link.springer.com	35,553
search.proquest.com	34,534
search.ebscohost.com	30,024
openurl.ebscohost.com	29,280
journals.sagepub.com	24,321

CONCLUSION

Database usage analysis allows librarians to learn various aspects of how users access the library's e-resources. A simple question like how many times a database is visited within a period is

fundamental to justify the institution's expenditure. Without the usage analysis, it will be very difficult for a library to investigate a model of the user accessing e-resources based on which custom service could be developed. Compared to the vendor-provided data, transaction log analysis provides librarians access to first-hand data, allows cross-database comparison, and can be combined with information from other systems for advanced studies. In this article, the author presented a home-grown system used at the University of Manitoba Libraries to analyze EZProxy SPU logs. The author studied how licensed databases were visited within two years (September 1st, 2019 – August 31st, 2019). The author also used transaction log analysis to demonstrate how the database visits performed at different locations and by various user groups change with time.

Since the EZProxy SPU log is being automatically imported into the system daily, it allows the author to analyze the database usage activities within a longer period. Further studies could be focused on the information-seeking behaviors from a particular user group, as such information is available for off-campus visits. This article briefly discussed how users in the Faculty of Graduate Studies group used the e-resources. More studies could be performed in this area to investigate a model of user activities for particular user groups. Meanwhile, although the article demonstrated the roles Ethernet and Wi-Fi play when it comes to database visits from campus, the statistical usage is split into two groups by the fact that IP authentication is used for on-campus visits and ID authentication for off-campus visits. If IP authentication is replaced in the future, on-campus users and off-campus users could be evaluated together and a more comprehensive model could be developed.

Appendix: UserGroup Table

UserGroup table in the database. Groups 1-15 and 17-29 are used for undergraduate students. Group 16 is for graduate students. UofM faculty and staff are in Group 100.

GCode	GroupName
1	Faculty of Arts
2	Faculty of Science
3	Faculty of Engineering
4	English Language Centre (2)
5	College of Medicine
6	Faculty of Management
7	Faculty of Agric. and Food Sci.
8	Faculty of Education
9	Faculty of Architecture
10	Faculty of Law
11	College of Pharmacy
12	Faculty of Social Work
13	College of Nursing
15	School of Art

GCode	GroupName
16	Faculty of Graduate Studies
17	College of Dentistry
19	College of Rehabilitation Sci.
20	School of Dental Hygiene
21	Faculty of Music
22	Faculty of Kines. & Rec Mgmt.
24	School of Agriculture
25	College of Medicine - PGME
26	Extended Education
27	University 1
28	Faculty of Environment, Earth & Resources
29	Faculty of Health Sciences
100	UofM Faculty and Staff
101	USB Faculty and Staff
200	Hospital Staff
201	MRHA
300	Distance Education

References

- Baker, G., & Read, E. J. (2008). Vendor-supplied usage data for electronic resources: a survey of academic libraries. *Learned Publishing*, 21(1), 48–57.
<https://doi.org/10.1087/095315108X247276>
- Blecic, D., Fiscella, J., & Wiberley, S. (2001). The measurement of use of Web-based information resources: An early look at vendor-supplied data. *College & Research Libraries*, 62(5), 434–453. <https://doi.org/10.5860/crl.62.5.434>
- Blecic, D., Fiscella, J., & Wiberley, S. (2007). Measurement of Use of Electronic Resources: Advances in Use Statistics and Innovations in Resource Functionality. *College & Research Libraries*, 68(1), 26–44. <https://doi.org/10.5860/crl.68.1.26>
- Coombs, K. A. (2005). Lessons learned from analyzing library database usage data. *Library Hi Tech*, 23(4), 598–609. <https://doi.org/10.1108/07378830510636373>
- COUNTER. (2019). What is COUNTER? Retrieved July 25, 2019, from <https://www.projectcounter.org/about/>
- Duy, J., & Vaughan, L. (2003). Usage data for electronic resources: A comparison between locally collected and vendor-provided statistics. *Journal of Academic Librarianship*, 29(1), 16–22. [https://doi.org/10.1016/S0099-1333\(02\)00400-7](https://doi.org/10.1016/S0099-1333(02)00400-7)
- Gonzales, B. M. (2018). Analyzing EZproxy SPU Logs Using Python Data Analysis Tools. *Code4Lib Journal*, (42).
- Liu, G., & Fu, P. (2018). Shared Next Generation ILSs and Academic Library Consortia: Trends, Opportunities and Challenges. *International Journal of Librarianship (IJoL)*, 3(2), 53–71. <https://doi.org/10.23974/ijol.2018.vol3.2.94>
- NISO. (2019). SUSHI History/Origins. Retrieved July 25, 2019, from

- <https://www.niso.org/standards-committees/sushi/historyorigins>
- OCLC. (2018a). LogSPU. Retrieved from https://help.oclc.org/Library_Management/EZproxy/Configure_resources/LogSPU
- OCLC. (2018b). Starting point URLs and config.txt.
- OCLC. (2019). Log files overview. Retrieved July 24, 2019, from https://help.oclc.org/Library_Management/EZproxy/Manage_EZproxy/Log_files_overview
- OED, O. (2019). database, n. Retrieved from www.oed.com/view/Entry/47411
- Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, 11(2), 41–66. <https://doi.org/10.1108/eb047884>
- Shibboleth. (2019). Shibboleth. Retrieved July 25, 2019, from <https://www.shibboleth.net>
- Wan, G., & Liu, Z. (2010). Knowing your users: The value of article database usage analysis. *Learned Publishing*, 23(3), 225–236. <https://doi.org/10.1087/20100305>

About the author

Wei Xuan is Assistant Librarian and Head of Libraries Systems at the University of Manitoba Libraries. He has been playing an active role in numerous projects focused on systems interoperability, data protection and accessibility. His current research interests include resource evaluation, data management and library technology.