Big Data Analytics and Electronic Resource Usage in Academic Libraries: A Case Study of a Private University in Kenya

Wakahia, Samuel Kairigo (MLIS)
Assistant Librarian, Pan Africa Christian University, Kenya

Abstract

The purpose of the study was to apply Big Data analytics as a tool for evaluating electronic resources usage in the academic library setup in Kenya with reference to the library of one private university. Log files of postgraduate students were mined from the server where the offsite access platform (ezproxy) has been installed. Descriptive statistical techniques such as mean, standard deviation and percentages were computed. Data was transferred to the Statistical Package for Social Sciences (SPSS) software was aided in the analysis. Results revealed that in terms of usage intensity, total URL count was 2,352, the highest user made 283 downloads and the mean URL count was 49 downloads. Further findings revealed that no user utilized more than 5 databases over a period of one year. The mean usage intensity score for respondents who were trained or orientated on e-resource usage was above average at 69.0 while those who had not received training were below average at 29.8. It was concluded that big data analytics is a necessary and powerful tool for investigating electronic resources seeking and usage trends and patterns within Kenyan university libraries. Through big data analysis and data mining, usage patterns and trends such as usage intensity that might not have accurately been revealed through other tools are unearthed. Big data analytics has revealed user preferences and intensity of utilization of various databases and helped in detection of redundant databases. From the usage patterns, it was clear that the level of utilization of the University library's e-resource platform was very low. Most of the databases accessible through the platform were redundant. Further, only two databases namely e-book central and ebscohost were the most popular among users while the rest were barely being utilized if at all. For most students, just one or two databases were sufficient in meeting their research needs. An integrated data analytics model for investigating university library's e-resources usage is proposed.

Introduction

With the growth of online business, there has been an upsurge of analytics to provide insights about the business environment that inform business managers on what decisions they should make concerning the products of their business entities. Analytics are therefore as old as online business. Big data analytics however is a fairly new concept especially to the library world (Teets & Goldner, 2013). Websites or web portals have since been developed which have a definite impact on the approach taken by libraries to organize, offer and facilitate the access to information resources and services. More and more digital content continues to be published on online platforms which make modern libraries and information centers different from what they were in the past. Technology then advances so rapidly, bringing along new developments in the research world within very short spans. Such technological advancements prompt the library institutions to move their resources and services online, hence making websites a vital component of library services, in their role of linking the various online resources and databases to enhance easier access among the patrons who intend to utilize them.

Librarians invest a lot of time and money resources in enhancing access to online database resources that form part of the library collection (Imler & Eichelberger, 2011). In order to justify the importance of these resources and to ensure that they are being spent well, they conduct user studies in order to obtain usage statistics. The studies may be in form of focus groups, surveys through the use of a black book, direct observation and questionnaire which are sent out and issued to the users in search for their responses and feedback. In most cases the statistics are in form of a tally sheet in paper form used to generate a report meant for the management who need a constant update of how the library users interact with the database resources at their disposal (Farney & McHale, 2014). However, the above methods do not bring out the best picture of how these online resources and databases are used. Imler and Eichelberger (2011) have stated that these methods may not bring out the exact impression on how these database resources are used by the library institutions' patrons. For instance, direct observation affects the behaviors of the patrons, since their physical presence brings out the psychological notion that they are being monitored. The focus groups, surveys and questionnaires just provide statistics but never bring out exact interactive picture between the patrons and the resources available.

While such methods are continually considered as noteworthy sources of data collection, they have notable limitations on their success levels in

bringing out the unbiased impact of utilization of library e-collection on any academic undertaking of an information institution. The only way to bring out the exact impression as well as unbiased results on how resources are utilized, especially the electronic resources and to help the information provider to fully dissect the needs of his information seekers is through data mining from big data collection. Renaud et al (2015) opine that data mining normally will entail the application of numerous data sets or even one sizable data set coupled with analytical techniques that will reveal trends or patterns. This research therefore sought to analyze and bring out the applicability of big data analytics as tools for understanding electronic resources seeking and usage behavior within one Kenyan private university.

Problem Statement

Big data analytics is an aspect that has been practiced for as long as online business has been in existence. Research in this area has been conducted in fundamental areas such as the business world, with an intention to understand the consumers and their ever growing needs, the political world, with an intention to understand the thoughts of the political electorate and the mobile phone industry where the main objective is to crack down on insecurity related issues. Within the library world, both globally and locally within Kenya, there is minimal and in some cases no attempt whatsoever to directly investigate how library users interact with the e-resources at hand through the application of big data.

With the library and research institutions spending money in millions enhancing access to the online resources that form part of the library collection, there is need to engage a tool that will present such details more accurately; Statistics show that in the United States, even with a varied nature of publisher prices being pushed down to vendors for specific periodicals, university libraries on average have to spend $172, $572 and $1,636 on single humanity journals, single social sciences and single pure science journals respectively in terms of annual subscriptions (Rose-Wiles, 2011).

In Kenya, Universities on average spend over one million Kenya shillings on journal subscriptions annually (Chibini, 2011). Such institutions have to justify the importance of these resources by ensuring that they are being spent well. Engaging data usage analytics for available databases and websites by studying the interaction patterns between the users and these resources will provide valuable insight into the effectiveness, relevance and the impact these resources have towards fulfilling the constant needs of the users. Such valuable insights will provide the management with concrete data meant for them to make informed decisions about issues like the subscription costs incurred annually paying for the e-resource databases. The same data will influence policy making among the top management in regards to

enhancement, retention and weeding of databases from the collection, intensified information literacy training and development and training of staff to acquire the skills needed to conduct the analytics. Skilled on board staff who can conduct the analytics would translate to timely data for quick decision making and presentation of data to the management in the required formats.

Research Objective

The objective of this study was to investigate big data analytics as a tool for analyzing electronic resources usage in academic libraries in Kenya.

Literature Review

Data is an essential component in all operations in the modern world since it appears as a representation of something that is occurring or that has already occurred. Meaning may not necessarily be drawn at an instance when looking at data sets but human beings have a great capacity to create meaning and therefore build knowledge out of obtained data sets through processing, studying and eventually analyzing such data sets. Frederick (2016) opines that Libraries, as with all other human activities, have an interest in data since it is a requirement for making the technologies within the library setup to work. The main strategic reasons that compel libraries to conduct analytics are to enhance the experiences of library patrons, to demonstrate value for finances allocated and to bring out excellence in research (Showers, 2014; 140).

Library patrons also require data in order to find or locate the information needed when they have an information gap to fill. Frederick further states that research and academic libraries often should make accessible statistical and other important forms of data which would frequently be used for learning, teaching, supporting decision making as well as fulfilling research centered purposes. According to Intezari and Gressel (2017) in as much as the datasets generated from the big data, that is available to organizations may vary marginally, such volumes of data usually present more robust and valid results when organizations are faced with queries. The most important thing is to understand what type of analytics to apply where in order to obtain a desired range of results.

The library as a result, in its role as an information resourcing institution not only gets exposed to so much information but also generates a lot of data in the process. Such large amount of data is what is referred to as big data. The exposure to big amounts of data which may either be structured or unstructured results in the emergence of a state of disorganization within the libraries where there is so much information at the library's disposal, some of which may result in an information overload and not be of assistance to the users. In some cases, the users give their feedback through available platforms and they don't get an actual audience at the library because the libraries don't

have staff, such as data librarians who are specialized in mining data sets that could reflect the queries or the unspoken concerns emanating from the users. This necessitates the need for constantly conducting big data analytics within the library setup to get timely and reliable data that shows how well e-resource data bases are utilized. According to Vecchione et al. (2016), librarians and other researchers can therefore gain insights into user behavior through assessing their research patterns in order to establish what resources in particular are the most popular and to a greater extent the unpopular resources, so as to decide on whether to improve their visibility on the website for quick navigation, access and retrieval by the users or to remove them altogether from the collection.

With the continued advancement in technology, and the introduction of different modes of study in institutions of learning, there has been an increasing need to ensure that the library resources are not only accessible from the ordinary desktop computers but also through other gadgets which library patrons and users can connect to either from within the library or remotely. Most users will opt to use their personal gadgets to access the available resources and for this reason, the library has no option but to adjust to this trend accordingly. Tidal (2013) concurs with this statement by saying that technological trends have enabled gadgets such as Wi-Fi-enabled e-readers, computer tablets and smart phones to connect library users to the available online resources and the information institutions have had to adjust accordingly in order to meet the arising needs of such gadget or smart device users. In order to analyze information seeking behavior or e-resource database usage trends, libraries can adopt web analytics processes, such as Google analytics, web log analysis through analysing cookies and browser history files, mining virtual private network logs, the use of social media and also independent query platforms such as contact forms.

Farney and McHale (2014) conceptualize web analytics as a method of gathering and trying to construct meaning out of the virtual usage of a website by its users in order to understand how these users interact with this website. In library and information science, it is seen as a process of tracking the significance of web and online information material available to an information institution. The process is done under the umbrella of webometrics. In library and information science, deeper research about the web users can be done through web log analysis which is a branch of webometrics and web servers keep track and record all data about a user activity, in regards to the dates of access, times of access and content the particular user viewed. This data can be manipulated to suit the interests of the researcher intending to study the web users. Data collected in this case will include the serial occurrences of visited websites, pages viewed by a user, and the amount of time spent on each page visited.

The most appropriate point to conduct web analytics when trying to analyze library users are on the library web portals since such portals act as library gateways or library openings that lead the users to the research databases available to that library either through timely subscriptions or through open access. Vecchione et al. (2016) argue that such portals and gateways are created with an intention of saving the users a lot of navigation and time thus enhancing quicker and more efficient research and task completion through consolidating all the resources onto a single web platform. Farney and McHale (2014:4) also point out that "web analytics capacitates libraries to look further than just data collection but to use the facts presented out of conducting the analytics for tangible action." Google analytics and altmetric systems can be attached to this web portal, to establish web page visits, counts of visitations and download counts based on every individual. Such information would come in handy to inform the library management how the e-resources databases available have being utilized by providing accurate and timely statistics on who newly visits the website, who's a constant visitor, countable searches and also session lengths which the users spend on particular databases. Bashinsky (2015) informs clearly that poor visits to the various sites could be as a result of outdated designs, reading and navigation difficulty of outdated plugins or slowness when it comes to loading. All these aspects can be insightful to prompt library management to take necessary action.

Libraries can also mine virtual private network logs to obtain usage data statistics. Virtual private networks provide access to local area network connectivity for people who are accessing such a network remotely. Libraries use virtual private network platforms such as ezproxy to provide off campus access to resources subscribed to the parent institution under a certain IP address. Magnuson and Davis (2014) state that "in as much as mining log data from a system such as ezproxy may look like an overwhelming talk, it can be extremely revealing to show the top databases visited by the users, what connection are done within a location or remotely, which users between the students and faculty are using what particular resources and even reveal the busiest days as well as when there is minimal activity." This means that through collecting of log data obtained through from the Virtual Private Network (VPN) platform, the library is able to understand the information seeking patterns from the users through studying and analyzing the user logs. The VPN systems have a log format that reveals the actual details through initials that represent the actual information needed. The statistics collected become very useful for conducting database cross comparisons and therefore inform on which particular databases are used the most compared to the ones that are minimally or hardly used. Such statistics will give a detailed account

of the intensity in usage therefore informing which databases are most valuable to the users.

Apart from direct data mining on proxy servers, in order to clearly understand how users utilize the databases at their disposal; the libraries consider avenues that enhance the building of rapport with the users. This helps the libraries to get feedback about the services on offer and even insights from the consumers about what they need most in order to reap utmost benefit. Such avenues include social media platforms created for the library set up with an intention to constantly communicate with the users and also get user feedback. McCallum (2015) argues that libraries can exploit the potential benefits of social media to build intact relationships with their patrons regardless of their location and the modes that they adopt for learning and accessing the library resources and services at their disposal. Commentary on published works done by scholars has expanded into the online world in the form of blogs, e-mails, tweets, web pages, and a host of social media opportunities for communication. Social media platforms would not only be used as a way of marketing information resources available, through recommendations and tags but commentaries and recommendations in this case go on to inform about the strength of the resource base within the library (Xie and Stevenson, 2014). When rapport is created through these social media platforms, the users can make comments about a poor service or applaud the library for a good service through the comments made through the same platforms. This comes in handy when the library is trying to gauge the satisfaction rate of the users in response to the services offered. Other rapport avenues are contact forms where users can also make requests through filling a contact form. Contact forms are a great way of encouraging consumers, both current and potential to take appropriate action when faced with a situation. He however argues that they should be kept easy and short to encourage potential clients to complete. Contact forms differ with social media platforms since the same way a user will write a long comment requesting for something is handled differently on a contact form since the contact forms will require information filled in brief.

Top research institutions and business organizations have invested heavily in technologies that best utilize the continually increasing opportunities for new data collection, as well as merging data from both internal and external sources and eventually bringing into application data analytics concepts to surpass the competitors' levels of production (Marshall et al., 2015). This new data normally regarded as Big data when managed effectively in its huge volumes provides opportunity for organizations to conduct a real time analysis based on what is depicted from this huge data volumes and also for the organizations to react accordingly at an appropriate time frame and within suitable deadlines. (Hurwitz et al, 2013).

Big data is characterized by huge sets of data whose size surpasses the typical databases or ordinary computers' processing capability. Through Big data, organizations realize a synthesized progression achievable when they obtain, analyze and leverage the information for the effective running of their daily operations. Big data therefore results in the evolution of the business organizations since through the obtaining, analyzing and leveraging the information gathered, the organizations are compelled to revise what they consider as initial business traditions. Through the re-evaluation of data that is available from without and within the confines of an organization, valuable insights are provided from such information since it highlights pertinent issues at a broader scope which when well utilized will lead to enhanced decision making processes (Runciman & Gordon, 2014). Cheng et al. (2016) opine that the more data is obtained; the structure of a problem becomes more illuminated and therefore can be observed more clearly. This will in turn lead to an accurate analysis drawn from the data to put in place a solution to the problem. Cheng further argues that data sets needed for the conduct of big data analytics cannot be handled by a typical relational or object-oriented database, ordinary computers, or the ordinary desktop application software since they are too huge. The need for huge parallel processing power of computer clusters becomes necessary therefore to be able to handle the massive data. Big Data Analytics, despite being a common practice within information and communication technology (ICT) circles, is new and at an infancy stage when it comes to library management application (McAfee & Brynjolfsson, 2012).

Information managers or librarians can use big data analytics applications to obtain valuable information about areas concerning service provision within the library and information centers to investigate information seeking trends when they interact with the e-resource databases and understanding what users utilize most. The investigation can also point out what areas are points of concern because they probably are not working for the library users. Insightful and valuable information from all the big data sources that a library is exposed to and that can be used for timely decision making can be obtained through the data mining process. Data mining in education and e-learning set up can be used in to obtain information concerning multiple problems. The process may come in handy when establishing problem relating to students' learning assessment, materials used for learning and evaluation of courses. It can also be used to guide on course adaptation based on learning behavior of the students. Simple distributed algorithms can be applied for mining and harvesting data not only from server platforms but also from the text sources such as e-mails as well as the internet.

Some authors have felt that the data mining concept may be an intrusion into people's lives. Farney and McHale (2014) argue that analytics tools through the data mining process have the ability to track a computer's IP

address, which is not only a sensitive issue when such an address is used to spot the network, Internet service provider (ISP), and estimated geographic location of the computer that uses a website, but it is also very invasive. Another aspect that can be looked at as a short coming for data mining through the use of logs is that such logs never reveal the actual motivation behind information searches conducted or even the step by step actions that the users engage in during the search. According to Chen et al. (2015) projected statistics based on log analysis alone can lead to people drawing very basic conclusions about information seeking behavior with no appreciation of the context under which the information search is being conducted or even the preferences of the particular users. Nevertheless Banse et al. (2012) are quick to opine that tracking cookies, are a crucial aspect that allow web content providers to link multiple sessions of a frequent visitor to a website and are used on many popular websites today.

Research Methodology

The study adopted a descriptive research design. Every user who needs to access the electronic resources has to have authentication through the ezproxy access platform. With permission from the case study university, log files of postgraduate students were mined from the server where the offsite access platform (ezproxy) was installed. in order to obtain individual user log files that register once a user logs into the ezproxy platform to access the electronic resource databases. The researcher sought data for a period not exceeding and not less than the past one academic year. Descriptive statistical techniques such as mean, standard deviation and percentages were computed. Data analysis was performed using Statistical Package for Social Sciences (SPSS) software.

Results and Discussions

Data mined from the library's server revealing URL count is presented in Figure 1. The figure shows that the mean URL count was 49, with individual counts per user deviating from this mean by about 63 URLs (M=49, SD=62.991). The results also revealed that the total URL count was 2,352 and the highest user made 283 downloads over the one year period. Out of the total URL count, the results showed that 50% of the users accounted for only 5.4% of the download volume. This is also reflected in the figure which depicts the skewness tailing to the right, suggesting that majority of the students were light users of the university library's e-resources. It was also established that the bottom third (33.3%) of the respondents had zero logs, meaning they never utilized electronic library resources over the one year period under review. In comparison, the top third (33.3%) of the users accounted for 78.9% (1855) of the volume of e-resources used as indicated by URL count while the remaining

third (33.3%) of the users shared 21.1% of the volume. The results suggest that just a third of the students did make adequate use of the e-resource materials availed by the university library. This analysis of the user log files has therefore enabled the determination of user intensity of the university's e-resources. This is in line with the argument fronted by Gooding (2016) that in library and information science, deeper research about the web users can be done through web log analysis as it keeps track and record all data about a user activity. This means that this element of Big data analytics is applicable in making sense of e-resource usage patterns in academic libraries.
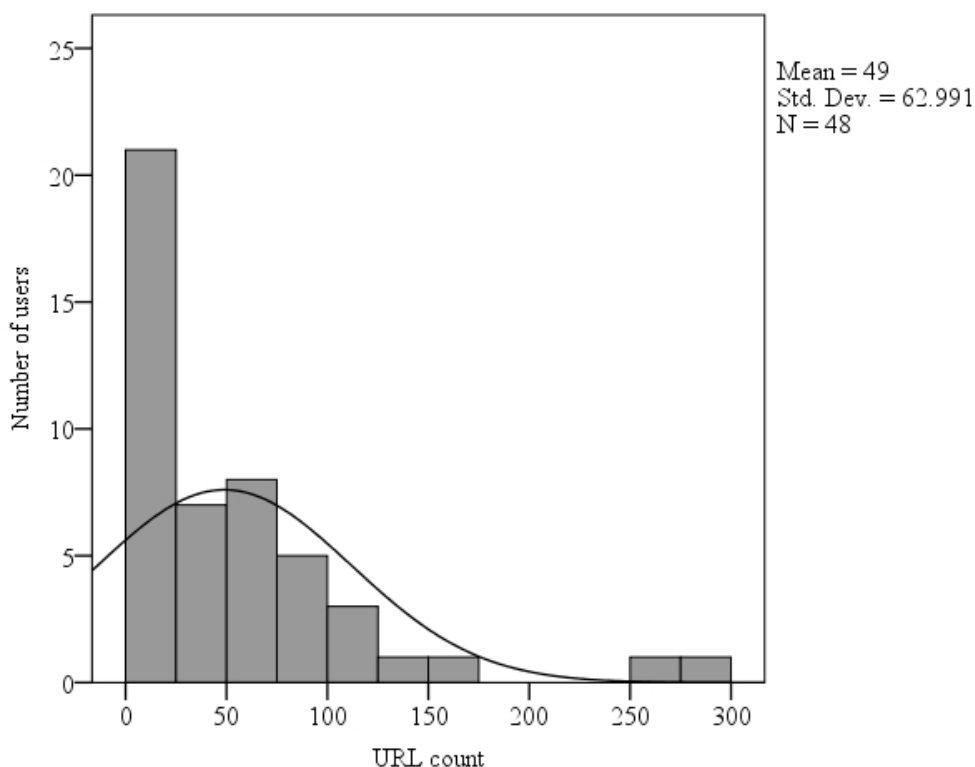

Figure 1 Usage intensity as measured by URL count

The size of user log files were also used as a proxy for usage intensity. Figure 2 presents the results. The figure shows that the average Log File size was 9 kilobytes, with a variation of 11 kilobytes (M=9.29, SD=10.892). It was also found that the total size utilized by students over the one year period was only 446 kilobytes. The biggest log file size was 48 kilobytes. Two-thirds (66.7%) of the users' log files collectively accounted for only 25% of the used capacity, of which some 31.3% (15) log files had zero kilobytes, implying no data on them. The results suggest that despite the intensity of research that characterize studies at postgraduate level, about a third of the students never utilized the university library's e-resources at all. Therefore, through big data

analytics, the data has revealed unbiased results on how resources are utilized, which helps information providers evaluate the needs of information seekers. This is consistent with the perspectives of Renaud et al. (2015) who associate big data analytics with depiction of the exact impression of usage levels of e-resources in universty libraries.
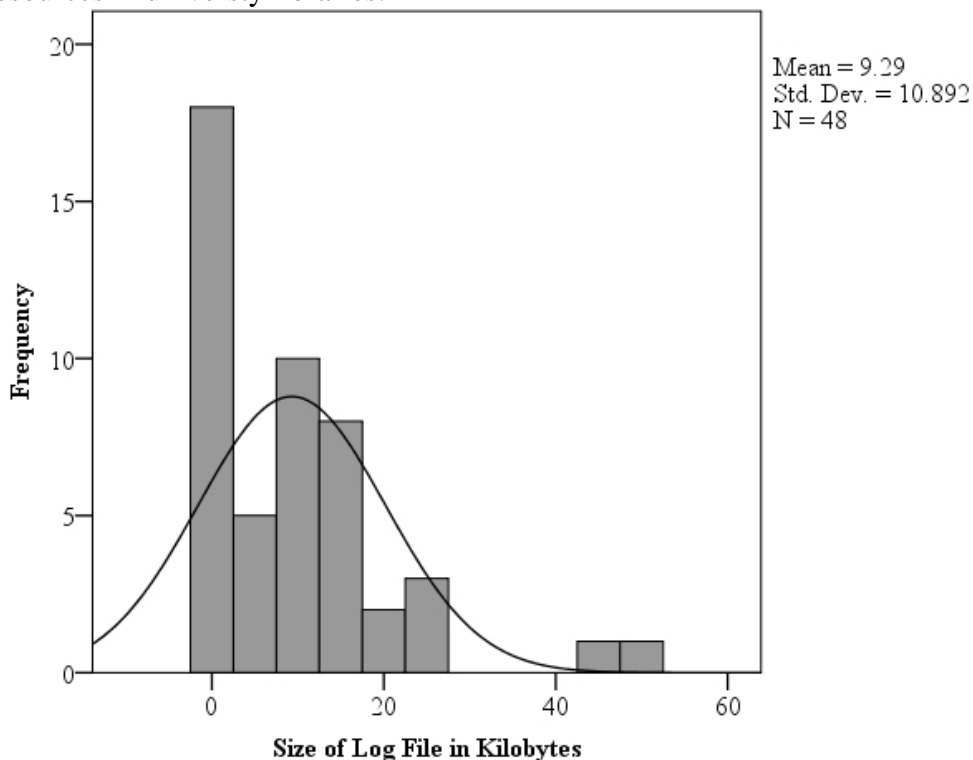

Mean = 9.29
Std. Dev. = 10.892
N = 48

Figure 2: Usage intensity as measured by Size of Log files in Kilobytes

The user log files were analyzed to establish user intensity as measured by database count. Table 2 shows that 64.5% (20) of the users accessed only one database, 32.3% (10) of the users accessed two data bases and only 3.1%(1) of the users utilized 3 databases. No user utilized more than 3 databases over the one year period under review. The results suggest that participants in the study did not make adequate use of the variety of databases accessible through the university library's e-resource portal.

Table 1 Usage intensity as measured by database count

Number of databases accessed	Frequency	Percent (%)
1	20	64.5
2	10	32.3
3	1	3.2
Total	31	100.0

The user log files were also analyzed to reveal usage patterns for various databases. Figure 3 ranks the databases by order of popularity. The figure reveals that the most popular databases were e-books (25.5%) and ebscohost (25.5%) collectively accounting for 51% of the e-resources used. These were followed by Jstor at 17.0%, Emerald at 12.8% and NTS library at 8.5%. The figure indicates that the top five databases controlled 89.4% of the share of e-resources utilized. The figure also indicates that some 4.3% of the users frequented Greenstone and Cambridge, each. The least popular data base utilized was Wiley Inter-science at 2.1%. None of the remaining 10 other databases availed by the University library namely Hinari, University of Chicago Journals, Doaj, Springer, Taylor & Francis, Mary Ann Liebert, Intech, Eldis, World Cat and APS Journals were accessed in the one year period under review. This indicates that majority of the e-resource databases were redundant to this particular group of users. The finding agrees with Magnuson and Davis (2014) who stated that system such as ezproxy can be extremely revealing to show the top databases visited by the users, which users utilize what particular resources and even reveal where there is minimal activity.
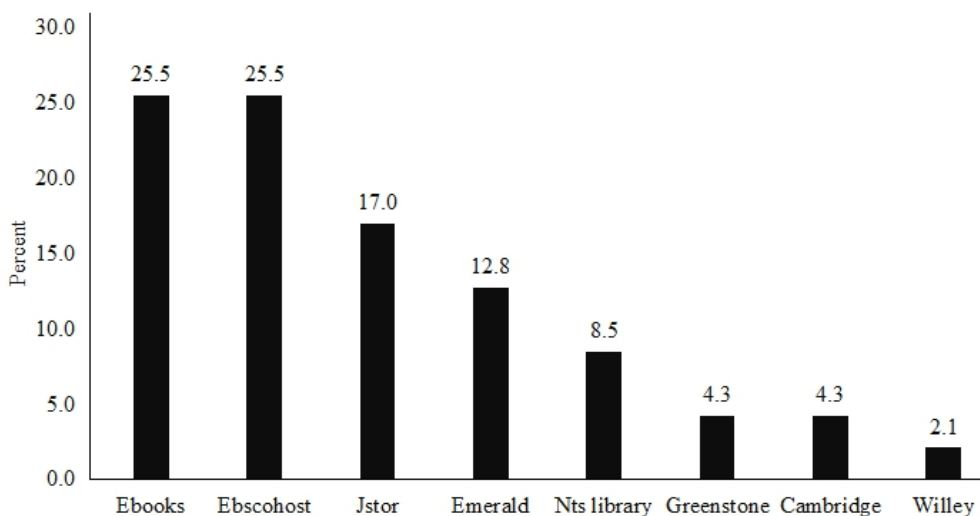

Figure 3 Usage patterns of specific databases

Conclusion

Big data analytics is a necessary and powerful tool for investigating electronic resources seeking and usage trends and patterns within Kenyan university libraries. Through big data analysis and data mining, usage patterns and trends such as usage intensity as measured by URL counts, log file size in kilobytes, database usage counts and specific databases accessed that might not have accurately been revealed through other tools are unearthed. It has

revealed user preferences and intensity of utilization of various databases and helped in detection of redundant databases. From the usage patterns, it was clear that the level of utilization of the University library's e-resource platform was very low. Most of the databases accessible through the platform were redundant. Further, only two databases namely e-book central and ebscohost were the most popular among users while the rest were barely being utilized if at all. For most students, just one or two databases were sufficient in meeting their research needs.

Overall efficiency of the university library's e-resources could be improved by removing redundant databases from the platform. If the university seeks to reduce costs, the ten databases that were never used over the one year period under review are attractive targets for cost-cutting. However, given the future potential for their use, such efficiency measures should take into consideration the long-term view and strategic direction of the university. A cost-benefit analysis of the same is recommended before implementing such a measure.

All the data mining and analytics technologies for e-resources usage were appropriate for this particular Kenyan university libraries except Google analytics and cookies which were irrelevant for tracking and analyzing e-resource usage in this research. Specifically, proxy server logs enabled the analysis of e-resource usage intensity and database preference that could not have been accurately revealed through any other technique. However, where online survey systems are not integrated in the library's e-resource platform, the power of big data can be enhanced greatly by the integration of traditional tools of data gathering and analysis such as off-line questionnaires and interviews. In the current study for instance, big data did not reveal the feelings, perceptions and attitude towards the university library's e-resource platform. In view of the study findings, an integrated data analytics model for investigating university library's e-resources usage is proposed as illustrated in Figure 4.
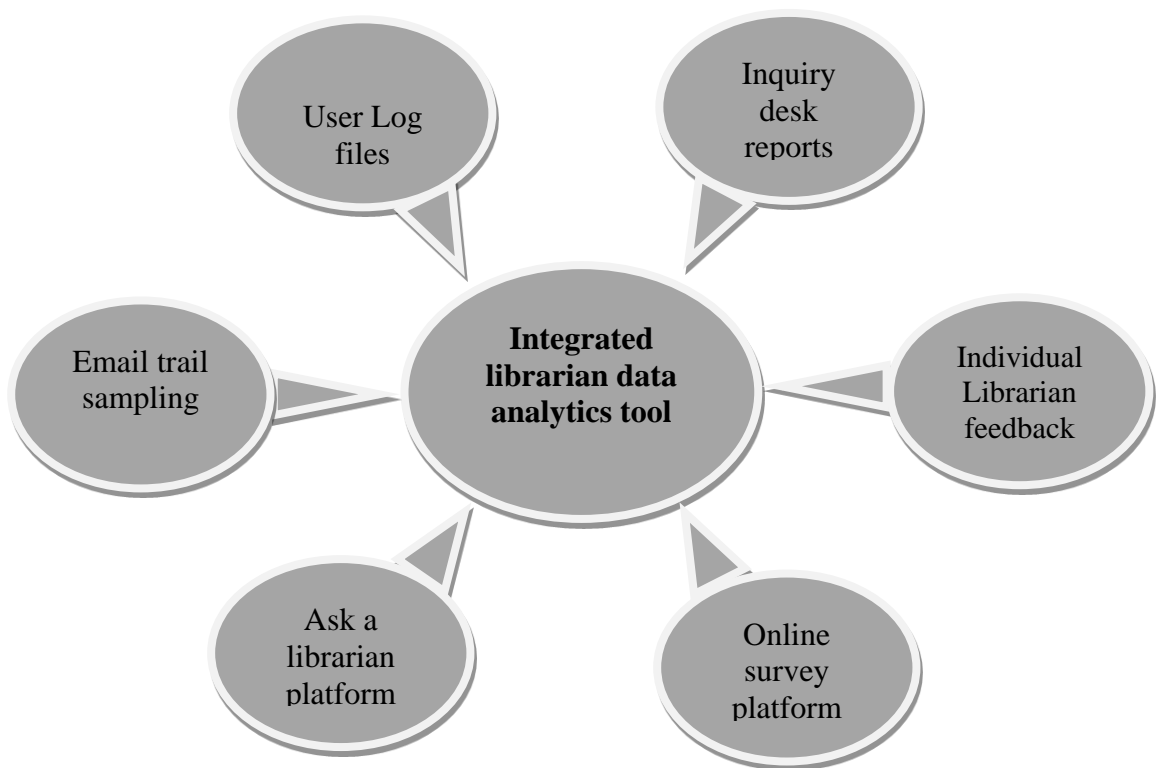
Figure 1 Integrated Data Analytics Model

The model in Figure 4 ensures that libraries are able to mine data from all potential fronts rather than limiting data analysis to information from log-files alone. This affords the university with the capability of correlating usage trends and patterns with subtle information gleaned from both online and off-line sources. The implication of this is that an online questionnaire should be designed and embedded on the university library's website. Similarly, feedback from interactive platforms such as email trails and ask-a-librarian platforms should be linked to the online survey questionnaire and user log files. Inquiry desk reports need to be compiled by librarian on duty on a regular basis. Of particular note is the importance of exit surveys by graduating students. In addition, an internal librarian feedback platform should be designed to open a feedback front for individual library staff to comment. Future research efforts is recommended on how the proposed integrated data analytics model can be developed into a software tool that generates real time reports to inform decision making by university managements. Such a software tool should have commercial value.

References:
1. Bashinsky, A., (2015) *16 reasons why people leave your website.* Retrieved on 27 March 2017from https://blog.hubspot.com/marketing/why-people-leave-your-website#sm.00011gi7o116bbfc7y3zeny3f63xt.
2. Cheng, S., Zhang, Q. & Qin, Q. (2016). Big data analytics with swarm intelligence. *Industrial Management & Data Systems*, *116*(4), 646-666.
3. Chibini, V. M. (2011). *Use of electronic journals by academic staff at Strathmore University*. Retrieved on 10th March 2017 from https://su-plus.strathmore.edu/bitstream/handle/11071/1566/Thesis%20Chibini%20Final.pdf?sequence=1
4. Farney, T., & McHale, N. (2014). *Library Technology Reports : Web Analytics Strategies for Information Professionals*. New York, NY: ALA TechSource.
5. Frederick, D. E. (2016). Data, open science and libraries – the data deluge column. *Library Hi Tech News*, *33*(8), 11-16.
6. Hurwitz, J., Nugent, A., & Halper, F. (2013). *Big data for dummies*. Retieved on 20th April 2017 from http://www.ebrary.com
7. Imler, B., & Eichelberger, M. (2011). Using screen capture to study user research behavior. *Library Hi Tech*, *29*(3), 446–454.
8. Intezari, A. & Gressel, S. (2017). Information and reformation in KM systems: big data and strategic decision-making. *Journal of Knowledge Management, 21*(1), 71-91.
9. Magnuson, L., & Davis, R. C. (2014). *Analyzing EZProxy Logs*. Retrieved on 15th March 2018 from http://acrl. ala. org/techconnect.
10. Marshall A., Mueck, S. & Shockley R. (2015). How leading organizations use big data and analytics to innovate. *Strategy & Leadership, 43*(5), 32-39.
11. McAfee, A. & Brynjolfsson, E. (2012). Big Data: the management revolution. *Harvard Business Review, 90*(10), 60-66.
12. McCallum, I. (2015). Use of social media by the library: Current practices and future opportunities. *Australian Library Journal*, *64*(2), 161-162.
13. Renaud, J., Britton, S., Wang, D., & Ogihara, M. (2015). Mining library and university data to understand library use patterns. *The Electronic Library*, *33*(3), 355–372.
14. Rose-Wiles, L. M. (2011). The high cost of science journals: A case study and discussion. *Journal of Electronic Resources Librarianship*, *23*(3), 219–241.

15. Runciman, B. & Gordon, K. (2014). *Big data: Opportunities and challenges.* Retrieved on 18th April 2017 from http://www.ebrary.com
16. Showers, B. (2014). Developing a shared analytics service for academic libraries. *Insights*, *27(2),* 139–146.
17. Teets, M., & Goldner, M. (2013). Libraries' role in curating and exposing big data. *Future Internet*, *5*(3), 429–438.
18. Tidal, J. (2013). Using web analytics for mobile interface development. *Journal of Web Librarianship*, *7*(4), 451–464.
19. Vecchione, A., Brown, D., Allen, E., & Baschnagel, A. (2016). Tracking user behavior with google analytics events on an academic library web site. *Journal of Web Librarianship*, 1–14.
20. Xie, I. & Stevenson, J. (2014). Social media application in digital libraries. *Online Information Review*, *38*(4), 502-523.