A log analysis study of 10 years of eBook consumption in academic library collections

Haley Littlewood, Annika Hinze, Nicholas Vanderschantz,
Claire Timpany, SallyJo Cunningham

Dept. of Computer Science, University of Waikato,
Private Bag 3105, Hamilton New Zealand
hml15@students.waikato.ac.nz,
{hinze,vtwoz,ctimapny,sallyjo}@waikato.ac.nz

Abstract. Even though libraries have been offering eBooks for more than a decade, very little is known about eBook access and consumption in academic library collections. This paper addresses this gap with a log analysis study of eBook access at the library of the University of Waikato. This in-depth analysis covers a period spanning 10 years of eBook use at this university. We draw conclusions about the use of eBooks at this institution and compare the results with other published studies of eBook usage at tertiary institutes.

1 Introduction

Most academic libraries are now offering eBooks and eJournals in addition to paper books and paper journals for today's researchers. For the purpose of this paper, an eBook is understood to be an electronic book rather than any electronic text. eBooks are usually viewed and accessed via specialist software or hardware such as eReaders and eCollections. The eCollections surveyed in this study are Ebrary and EBL. While the selection and consumption of physical books has been comprehensively studied (e.g, [14,30,31]), the library patrons' habits in accessing eBooks have been less well explored (e.g., [22] studied how patrons make relevancy decisions when selecting eBooks). In the current paper, we explore the consumption of eBooks, in particular the use of patron-driven acquisition, eBook acceptance and usage, and the handling of eBooks in an academic library. In a log-analysis study, we explored the use of eBooks at the library of the University of Waikato since 2003 when digital documents and eBooks began to be offered to students and staff.

2 Related Work

Understanding usage patterns of physical libraries, digital libraries and eBook systems can be of benefit to Librarians and researchers in Digital Libraries. Even though a number of studies on eBook collections have been reported [1,27,21,22], Tucker observes

that due to the rapidly and constantly evolving nature of this field, the conclusions from works published as recently as five years previously may or may not be applicable to the current environment [32]. Furthermore, the variation in data made available complicates comparative analyses [8,9,26,29]. Studies looking at book use in libraries and eBook collections seem to approach the issue from one of two directions. Observations and interviews have been conducted to better understand the decision-making process in the selection and consumption of eBooks (e.g., [30,31]). Transaction log analysis is also commonly used in studies to understand user behaviour and information seeking in online library catalogues [5,4] and digital libraries [19,24]. Using transaction log analysis can help to provide insight into aspects of the reading process that may not be easily observed otherwise [25, 22]. Previous studies using the Ebrary system have looked into user penetration as a tool for librarians improving their services, e.g., [26]. It was observed that managing multiple eCollections from different publishers can be problematic for librarians. Other studies have looked into subject-related acceptance of eBooks, and publisher market share, e.g., finding that eBooks in health sciences and hotel had highest usage [32]. Overall, the lack of recent data on eBook selection is widely acknowledged [28, 22, 14], and an awareness for the frequency with which patrons abandon their search [23] indicates a need for further research. The patrons' difficulties with digital systems for eBook search and selection likely result in lost sales (in online bookshops) or unwarranted paid use (in libraries, where readers perceive books as "free" [13]).

3 Methodology

We used data from the academic library of the University of Waikato, a small institution with about 12,550 undergraduates, 2800 postgraduates, and 628 academic staff.

3.1 The eBook collections

The University of Waikato (UoW) Library offers a variety of eBook platforms, access to which depends on the Library's licence agreements with individual providers. The eBooks studied here are provided by Ebrary. From 2003 to 2008, Ebrary was the only provider of eBooks to Waikato, with offerings increasing from an initial 30,000 titles in 2003, to 110,000 in 2008. Since 2008, the Ebrary offerings have remained at this level while other providers have been gradually included in the UoW catalogue, which provides access to 180,657 eBooks as of March 2014. Some eBooks are acquired as part of whole collections and others as single titles, according to need. The Ebrary offering, while static in overall numbers, remains dynamic with new titles being included, while older, unused eBooks are discontinued. All eBooks on offer are available via the library catalogue; most are owned by the library, with an increasing minority becoming available on an ad-hoc basis (patron-driven acquisition [22,12]). The difference between owned and un-owned eBooks is invisible to readers.

The Ebrary eBook holdings comprise a general academia-oriented collection, with a small number of supporting or recreational titles; eBooks are mainly categorized as

Business & Economics, Social Sciences, Science, Education, Technological Science and Medicine, while the least populated categories are specific, such as Gardening, Pets, and Bibles. The library also owns a large number of small, specialised collections aimed at specific subjects, and a small number of larger, more generalised collections that mostly contain journals, but may also contain eBooks. Librarians may buy single books and access to small specialised collections. They may also maintain access to the "Academic Complete" collection within Ebrary, which changes periodically and titles are added and removed as a consequence. Within the academic collection, books may get deleted both individually and in groups by the publisher, independently of librarians. Librarians can change single-use into multi-use books, some of which may still be subject to restrictions on the number of simultaneous accesses. Some collections are removed as they are shut down, or merged into another collection.

3.2 eBook log analysis

The data on which this study was based was collected over the period of 1 August 2003 to 31 December 2013, covering 11 out of the almost 12 years that eBooks have been available at the UoW library. The data thus represents a range of academic users including students, academic staff and research students. The Ebrary logs are available in a form that is anonymised by individual readers. Unlike previous studies [23], we did not have access to data that would allow us to track user sessions. eBooks that were deleted at some point from the catalogue only appear in the statistics if the book was previously accessed. There are no statistics available about how many books have been deleted and when. The statistics for accessed books are complete, including deleted books. All statistics used for this paper concern Ebrary books only, on data from both the main library and the specialist Education library. Statistics for the numbers of eBook on offer are taken from the University's annual reports covering all eBook resources.

4   Log analysis of eBook accesses

We explored six aspects of eBook usage through the Waikato University logs: access over time, popular eBooks, the number of pages accessed or printed, single vs multiple access schemes, queuing for access and patron-driven acquisition.

4.1 eBook access over time

When analysing the eBook access over time (both number of user sessions and number of pages accessed) a clear pattern can be seen that correlates with the timing of semesters A (March to June), B (July to November), and summer school (January).
Figure 1 shows the number of pages viewed (blue bars) and the number of user sessions (red line). Pages viewed including all pages that were turned to, regardless of how long the user spent on the page. Since 2011, eBook usage has steadily increased for both A and B semesters as well as during down-times (such as between semesters and during

summer breaks). A comparison of yearly user sessions confirmed an increase from 1431 in 2003, 19252 in 2008, to 42,225 in 2013.

We next explored whether this increased usage was caused by an increase in the number of eBooks available or by changes in staff and student numbers. The average pages viewed per accessed book and month was found to remain relatively stable at between 10 and 20 pages. While the number of staff and students remained relatively stable between 16,000 (2003) and 14,000 (since 2005), the average number of accessed
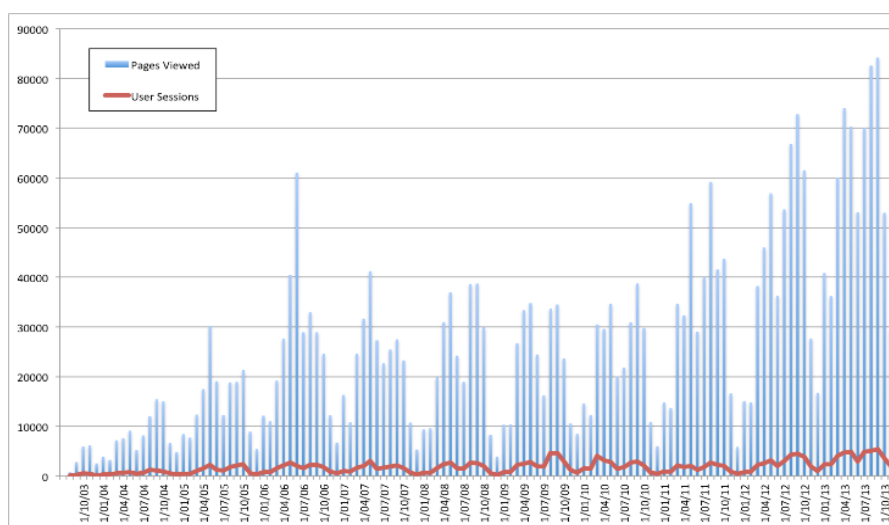

Figure 1. Number of user sessions and pages viewed over time

books per person was increasing from 0.5 (2005) to 1.4 (2009) and 2.2 (2013). We observe that over the years, an increasing proportion of books were accessed, from about 2% (2003) to about 10% (2005), falling to 7% over 4 years (2011), and steadily increasing since then to 14% in 2013. The observed dip was caused by disparity between stability in the numbers of accessed books while the overall number of books grew. Overall, there were 44,001 unique book accesses between 2003 and 2013. We note that between 2006 and 2011, the increasing numbers of books available did not seem to influence the number of books accessed. We conclude that the increase in number of pages and eBooks accessed is not caused by an increase in patrons or available eBooks, but by an increased use of eBooks over time. Of the books available at UoW library via Ebrary (13,899), about 39% have been accessed since 2003.

4.2 Popular eBooks

Of the most frequently accessed books, only five were found to be eReserve books, i.e., eBooks that had been used as the recommended text in a course. All of the eReserve books can be found among the top 38 of most accessed. While, on average, each accessed book was involved in 5.2 user sessions, the top 49 books have been accessed 180 to 1,300 times. The top category was Social Science, from which 10 eBooks were

accessed overall 3,720 times. The areas with the second highest usage were Education, and Business & Economics (8 books each). The areas with the highest average access per book were Philosophy (667 on 1 book) and Psychology (480.5 on 4 books).

4.3 Pages accessed/read/printed

Figure 2 shows the number of pages that were accessed online or printed out, ordered by subject area. Though 10 of the 49 top accessed books were from Social Science, this area comes second in the number of pages accessed (370,000 pages). Business & Economics (8 books in the top 49) is first with more than 500,000 accessed pages. The explanation for this seeming anomaly might be either that business books contain more pages than social science books or that business students and researchers read more pages of the eBooks they access. In comparison to books available in each field (Fig.2), we note that for most subjects the number of pages viewed rises with the number of books available (e.g., most available books are from Business & Economics, and so are most pages viewed). Exceptions are Political Science (proportionately fewer pages viewed), and Education (proportionately more pages viewed). This might be due to
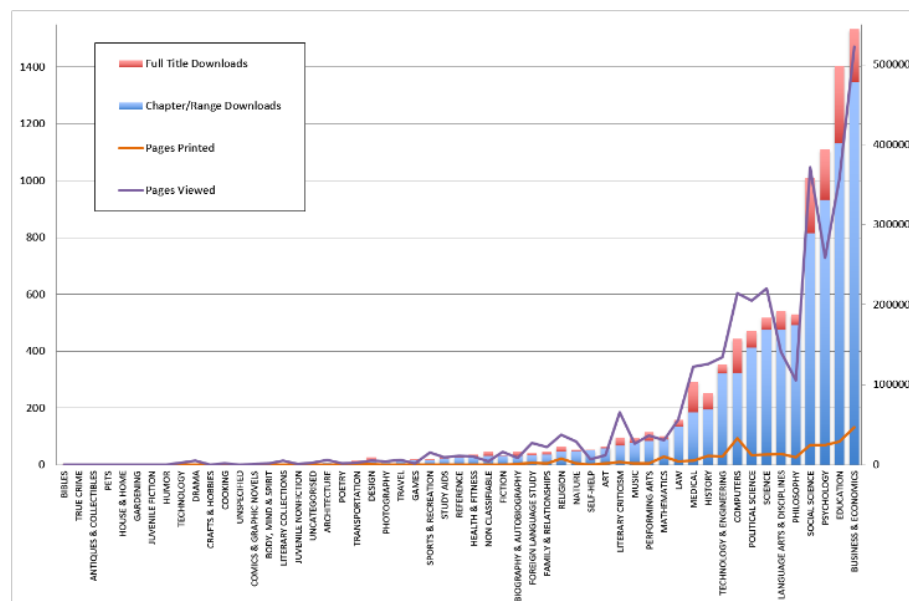

Figure 2. Pages accessed and pages printed downloads (range vs whole books)

either different patron behaviour or different book lengths in different fields. There are a number of areas in which large proportions of eBook pages are printed for reading (orange in Fig. 2). The classifications that show the largest proportion of printed to viewed pages are: Antiques & Collectibles (162% -- more pages printed than viewed), Bibles (70%), and Humour (45%). These most likely refer to private interests rather than study and research purposes. Fields with more than 20% printed pages are Mathematics (35%), Transportation (28%), Gardening (26%) and Religion (22%). It appears

that Computer Science (18%) and Mathematics (35%) are the two subjects where patrons print significant proportions of accessed pages. Downloads of eBooks in part or in full has only been available since January 2012; it varied between 35 and 120 per month in Ebrary.

4.4 Single user vs multiple user books: Queues and Turnaways

The eBooks available through the UoW library are predominantly single user books (SUPO) and multiple user books (MUPO). Both of these hold perpetual access licenses. Other access schemes are short-term *loans* for one day or for seven days, and patron-driven acquisition (leading to perpetual licenses). Attempting to access a single user eBook that is already in use may mean a patron has to wait. *Queues* refer to users waiting for access to an eBook with single-user access rights that some other patron is currently using.

Figure 3. Wait Queues per Primary Category

patron is currently using. If a user accepts to join a wait queue to access a book, they are shown a window giving their position in the queue. Once the user reaches the front of the queue they are prompted again. They are then given 5 minutes to access the document before access is revoked and given to the next person. If they choose to open the document within those 5 minutes then they have unlimited access (until they close the document), unless there are other users queued. Figure 3 shows the number of users that had to wait for a book to become available before they could access it, sorted by category. The data refers to all users that chose to join a wait queue, regardless of whether they reached the book or left early. *Turnaways* refer to users who, when asked in a session to wait for an eBook to become available, decided not to enter a queue. Since August 2010, there were 508 turnaways in total for eBooks at the UoW library.
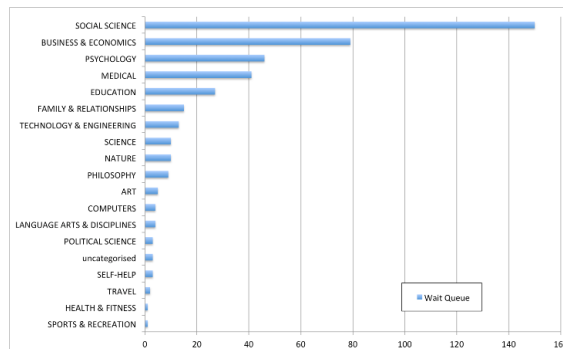
4.5 Patron-driven acquisition

In 2010, the UoW library started offering patron-driven acquisition (PDA) of eBooks, in which a patron selecting an eBook from the catalogue triggers the acquisition of this eBook. Statistics for PDAs were available from April 2012 onwards, and detailed acquisition data since July 2010. Figure 4 shows a comparison between sessions accessing librarian-acquired books (blue) and sessions on patron-driven acquisi-
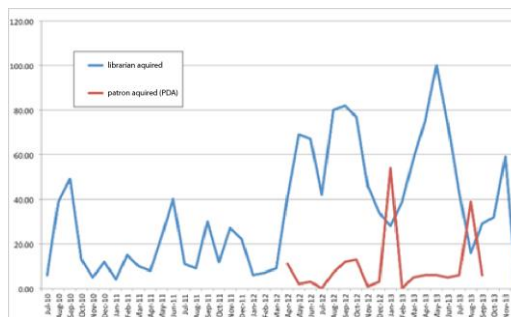

Figure 4. Access of eBooks acquired by librarians vs patron-driven acquisition

tions (red). We observe that few PDAs are accessed/acquired each month, and moreover, that they do not follow the semester structure, i.e., PDA peaks are seen in January and August. One aspect of these anomalies can be explained by the funding structure: there is only a small budget set aside for pre-paid PDAs. This budget is typically exhausted relatively quickly and is not always renewed. This may account for the months with very few PDAs and also for the absence of PDAs between September and December 2013 (as there was most likely no budget left). From 04/2012 to 08/2013 (the period for which statistics are available), on average 150 to 250 pages were accessed per PDA book (outlier: 08/2013: 970 per book, 16 books). There are no wait queues associated with PDAs as these eBooks are bought immediately when requested. When there is no budget for PDAs these eBooks are not shown as being available. Overall, since July 2010 1,449 PDAs were triggered (during 179 sessions, no information about distribution). As users are not prompted to confirm PDAs, it is possible that through research in areas that are poorly served by available eBooks, a small group of patrons inadvertently triggered a large number of PDAs.

5   Discussion

Comparison with similar studies. Other academic studies also observed that eBook usage follows the semester structure [2,3,27]. The British national eBook project [17] found that text-book usage varied by more than 50% depending on the teaching and assessment calendar. For the University of Waikato, these fluctuations are much more pronounced with differences of up to 600% recorded between the summer break and the height of B semester in 2013. The additional mid-summer peak observed in the Waikato data (December/January) was not found in other studies and might indicate the beginning of a greater use of eBooks during summer school or for research and class preparation. Further investigation of course offerings and reading assignments for that period would be necessary for clarification. Similarly, an increase of eBook usage was widely observed [3,22,27,32]. Tucker et al report a 54% increase over 3 years, with a 26% increase in pages viewed [32]. The UoW data from a similar time period shows an even higher increase (both eBook and page access between 2010 and 2013 increased

more than 200%). Again, because of the large degree of uncertainty caused by different recording mechanisms, it is hard to draw clear conclusions from this data beyond identifying an increased use of eBooks. The British national survey found that eBooks were predominantly used for quick look-up of facts and brief viewing at Universities [17]. While we also found these same usage patterns (i.e., accessing a few pages only), the average page access per book was 10 to 20 pages and quite a number of books were downloaded or printed in their entirety. The highest numbers of printed pages were found in Mathematics and Computer Science, the reasons for which could not be determined from this log analysis, but would warrant further exploration. The percentage of all accessed eBooks out of all the collections on offer seems to vary greatly between universities: Waikato patrons have accessed in total about 39% since 2003 (approximately 10% each year), while Sprague and Hunter reported that 16% of Ebrary titles were assessed at least once for the University of Idaho [29], at Australia's Edith Cowan university, 8% of eBooks were browsed and 5% were read (over a three year study period) [1,2], and Swinburne patrons accessed about 48% and created loans for approximately 19% in one month alone (these percentages are calculated from the usage data presented in [22]). Lamothe found strong correlations between the size of eBook collections and the level of eBook usage [20]. We were unable to confirm this for the UoW collections, as the number of eBooks available did not seem to influence the number of eBook sessions. Lamothe found that of all factors examined during their study the size of the collection had the strongest influence. They suggested that both the size and the content of a collection determines patron acceptance and utilization. We would argue that certainly the quality of the content matters in eBook acceptance, but the size may not matter as much once a critical mass has been reached.

Furthermore, the difference in ease of access between the various types of eBooks may also heavily influence their usage. At UoW the eBooks that the library owns are indicated as part of search results. However, eBooks that are available to UoW patrons via PDA do not automatically appear in search results. The patron has to explicitly ask for these to be included.

While Sprague and Hunter found that fewer than 2% of their eBooks have had repeated usage [29], we observed that over half of the eBooks accessed each year had been accessed in a previous year. Even with slightly inflated usage results (due to multiple accesses by the same user being counted as separate sessions), this is a significant difference. We found that course reserve books are only partially responsible for this high degree of re-use. Further analysis is needed about re-use of eBooks.

Statistics and comparability. A noticeable hurdle in the evaluation of eBook usage is the data availability. For example, we were unable to evaluate Waikato usage for titles from publishers other than Ebrary. Even if data is available, it is often not directly comparable. This has repeatedly been mentioned in other publications about eBook usage analysis. Cox compared the statistical data available for different eBook vendors with those in COUNTER [6] and SUSHI [8,9]. Counter is an international, extendible Code of Practice for e-Resources that allows the usage of online information products and services to be measured in a consistent and compatible way. SUSHI is a SOAP web

services "wrapper" for the XML version of Counter reports. Counter lists 130 publishers as being compliant for journals and databases with only 34 compliant for eBooks, although there are more than 100 unique eBook publishers [8,9]. Furthermore, EBook publishers typically use hosting companies and aggregators as their route to market. We wish to record our agreement with both their observation that eBook usage data is essential for libraries and their opinion that standardisation efforts are "prizes well worth fighting for" [9]. The main finding of our study is that it is very difficult to draw reliable conclusions from the data currently available. Because previous studies used different statistics (as there was no standard available) and covered shorter time spans, we can detect usage patterns, but find it hard to verify if these are typical for this type of institution (e.g., much higher fluctuation between semester and break times than in other studies, greater increase in eBook usage over the last few years, etc.) Similar to Sprague and Hunter [32], we appreciate the data and patterns we found, but feel stymied in making sense of it in the wider context. Furthermore, the data available makes it almost impossible to identify the reasons for some of the observed patterns. There are no clear records kept as to what activities took place over the last 10 years to promote eBook usage. Thus it is impossible to say which of these activities might be responsible for the high uptake of eBooks in some areas. Because the data is anonymised, the true benefits of a longitudinal study (e.g., changes in the patrons' usage patterns) are not as conclusive and specific as they might be had this data been available. At present, reuse and revisits cannot be identified. Furthermore, by using log analysis we are restricted to the observable patterns without external verification. It would be interesting to compare our findings with insights from patron observations and interviews.

6 Conclusions

This paper provides a log analysis study that covers the 10 years since the University of Waikato library started offering access to eBooks. Our study covers a longer period of time than is typically reported in other similar studies and gives insights into the trends of eBook usage at an academic institution. We found that our log analysis almost raises more questions than it answers. For example, neither the reasons for printing so many books or pages, nor the level of eBook usage during the summer months can be explained satisfactorily. When analysing the data for the university, we may conclude that the library has been successful in promoting eBook use as we found a much higher uptake and usage than in other studies. However, we are unable to identify the exact activities that contributed to their successful introduction. We conclude that libraries need to keep clear records of their efforts to promote eBook usage as a retrospective analysis based on eBook numbers alone is not sufficient. From the data available it is also not clear whether the results from Waikato are in fact significantly different from those of other institutions (as the statistics are difficult to compare). We encountered a number of obstacles when comparing the data to other studies available. In order to eliminate these, we recommend the use of standardised log keeping for eBook access that provides for identification of user patterns (while anonymising patron-specific data).

7   References

1. Ahmad, P. (2013). E-book Adoption in Academic & Research Libraries: Self-Reported Information Behaviour.
2. Ahmad, P., & Brogan, M. (2012). Scholarly use of e-books in a virtual academic environment: A case study. *Australian Academic & Research Libraries*, *43*(3), 189-213.
3. Christianson, M., and Aucoin, M. (2005). *Electronic or print books: Which are used?* Libr. Collect. Acquis. 29(1), 71–81.
4. Cooper, M. D. (2001). *Usage patterns of a web-based library catalog*. In *Journal of the American Society for Information Science and Technology* 52, 2 137-48.
5. Cousins, S. A. (1992). *In their own words: An examination of catalogue users' subject queries*. J. Inf. Sci. 18, 5 329-41.
6. COUNTER (2007). COUNTER: Counting Online Usage of NeTworked Electronic Resources. www.projectcounter.org.
7. Cummings, I. (2011). *Kansas libraries lead way in e-book access*. Kansan.com. Kansas, USA.
8. Cox, J. (2008). Making sense of e-book usage data. *The Acquisitions Librarian*, *19*(3-4), 193-212.
9. Cox, L. (2011). Librarians' use of usage statistics for journals and e-books. *Learned Publishing*, *24*(2), 115-121.
10. Gibson, C., and Gibb, F. (2011). *An evaluation of second-generation eBook readers*, The Electronic Library, 29(3), pp.303–319.
11. Hamblen, Matt. (2011). *Amazon: E-books now outsell print books*. Computerworld http://www.computerworld.com/s/article/9216869/Amazon_E_books_now_outsell_print_books2011, (last accessed April 11, 2012).
12. Hardy, G., and Davies, T. (2007). *Letting the patrons choose: using EBL as a method for unmediated acquisition of ebook materials*. In *Information Online* (ALIA, Sydney, 2007).
13. Hernon, P., Hopper, R., Leach, M. R., Saunders, L. L. and Zhang, J. (2007). *E-book use by students*. The Journal of Academic Librarianship 33, 1 3-13.
14. Hinze, A., McKay, D., Timpany, C., Vanderschantz, N., and Cunningham, S.J. (2012). *Book selection behaviour in the physical library*. In JCDL 2012, Baltimore, MD.
15. Holt, K. (2010). *E-book sales statistics from BISG survey*. In *Publishing Perspectives*. 2011.
16. Jantz, R. C. (2001). E-Books and New Library Service Models: An Analysis of the Impact of EBook Technology on Academic Libraries.
17. Jamali, H. R., Nicholas, D., & Rowlands, I. (2009, January). Scholarly e-books: the views of 16,000 academics. In *Aslib Proceedings* (Vol. 61, No. 1, pp. 33-47).
18. Jansen, B. J. and Spink, A. (2006). *How are we searching the WWW? A comparison of nine search engine transaction logs*. Information Processing & Management 42, 1 248-63.
19. Jones, S., Cunningham, S. J. and McNab, R. (1998). *An analysis of usage of a digital library*. Proc. ECDL'98, Springer, 261-77.
20. Lamothe, A. R. "Factors influencing the usage of an electronic book collection: size of the e-book collection, the student population, and the faculty population." *College & research libraries* 74.1 (2013): 39-59.
21. Littman, J., & Connaway, L. S. (2004). A Circulation Analysis of Print Books and E-Books in an Academic Research Library. *Library Resources & Technical Services*, *48*(4), 256-262.
22. McKay, D., Hinze, A., Heese, R., Vanderschantz, N., Timpany, C. & Cunningham, S. J. (2012). An Exploration of ebook Selection Behavior in Academic Library Collections. Proc. TPDL '12, 13-24.

23. McKay, D., Buchanan, G., Vanderschantz, N., Timpany, C., Cunningham, S. J., & Hinze, A. (2012). Judging a book by its cover: interface elements that affect reader selection of ebooks. In *24th Australian Computer-Human Interaction Conference* ACM, 381-390

24. Mahoui, M. and Cunningham, S. J. (2001). *Search behavior in a research oriented digital library*. In *Proc. Of the 5th European Conference, ECDL* 2001, Springer, 13-24.

25. Marshall, C. C. (2010). *Reading and Writing the Electronic Book*. Chapel Hill, NC: Morgan & Claypool.

26. McLuckie, A. (2005). E-books in an academic library: Implementation at the ETH Library, Zurich. Electronic Library, 23 (1), 92–102.

27. Obradovic, K. (2006). *Intrepid Traveller: the University of Auckland Library on the E-Book Journey*. VALA - Victorian Library Association Conference, Melbourne Australia. (2006).

28. Rowlands, I., Nicholas, D., et al. (2007). *What do faculty and students really think about e-books?* Aslib. Proc. 59(6), 489–511.

29. Sprague, N., & Hunter, B. (2009). *Assessing eBooks: taking a closer look at eBook statistics*. Library Collections, Acquisitions & Technical Services., 32(3-4), 150-7.

30. Stelmaszewska, H., Blandford, A. (2004). *From physical to digital: a case study of computer scientists' behaviour in physical libraries*. JoDL 4(2), 82–92.

31. Stieve, T., Schoen, D. (2006). *Undergraduate Students' Book Selection: A Study of Factors in the Decision-Making Process*. J. Acad. Libr. 32(6), 599–608.

32. Tucker, J. C. (2012) *Ebook collection analysis: subject and publisher trends*, Collection Building, Vol. 31 Iss: 2, pp.40–47.