Thank you for downloading this document from the RMIT Research Repository.

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.


Citation:

See this record in the RMIT Research Repository at:

Version: Accepted Manuscript

Copyright Statement: © 2013 Australian Library & Information Association

Link to Published Version:
http://dx.doi.org/10.1080/00048623.2013.773866
Practical overlap – the possibility of replacing print books with e-books.

Craig Anderson*, Jeanie Pham

*Corresponding author: Craig Anderson, University Library, RMIT University, Melbourne 3001. Email: craig.anderson@rmit.edu.au

Jeanie Pham, jeanie.pham@gmail.com, Student, Master of Information Studies, University of Canberra

Keywords: e-books, electronic and book format comparisons, academic libraries, overlap studies, collection development

Abstract
The increasing availability and prominence of e-books has led to a perception in some quarters that the requirement for physical books is decreasing, as most, or all scholarly material is openly available in digital form. Some academic administrators and librarians have suggested that it may be possible or desirable to have a solely, or predominantly electronic academic library. This study tests that assumption by randomly selecting a sample of titles from a mid-size Australian academic library and searching for suitable electronic copies. The conclusion of the paper is that a maximum of 33% of titles held by the case study library are available as e-books, and on a practical basis about 26% of titles held by the case study library could be readily obtained as e-books.

Context
For a number of years, the demise of print-on-paper books and the rise of e-books has been enthusiastically reported or bemoaned from many quarters. The recent literature, particularly outside the information industry, has claimed that e-books are replacing, or will soon replace, paper based print books. Part of this has been driven by Amazon’s announcement in various popular media that e-books sales have surpassed print sales. This has been widely repeated by the media with various degrees of accuracy. In 2012, The Guardian reported: “Kindle e-book sales have overtaken Amazon print sales, says book seller. For every 100 hardback and paperback books it sells on its UK site, 114 e-books are downloaded in a ‘reading renaissance’” (Malik 2012).

Rise of e-books
The e-book has a rather long history for a seemingly recent idea. While some sources claim that e-books started as early as the 1940s with Roberto Buso’s Index Thomisticus, the current form of e-books commenced in the mid to late 1990s. E-book sales were relatively slow and limited to specialized markets such as libraries until the introduction of the Amazon Kindle in 2007. Since the introduction of the Kindle, media and public attention paid to e-books has increased greatly. Education industry reports such as The 2011 Horizon Report (Johnson et al. 2011, 8) confidently predicted that e-books would be adopted in their “One Year or Less” category some two years ago.
Impressions of e-book availability

There is a widespread impression amongst the general public that the majority of print-on-paper books are also available in electronic format. This impression appears to be held by some academic administrators as well; evidenced by the report “Redefining the Academic Library” (University Leadership Council 2011). This report is produced in association with The Education Advisory Board (www.eab.com) which is part of the Advisory Board Company. The Advisory Board Company “Is a global research, technology, and consulting firm partnering with 150,000 leaders in 3,700+ organizations across health care and higher education” (http://www.advisory.com/About-Us). The Education Advisory Board is supported by, and works on behalf of an audience of “provosts, student affairs executives, chief university administration and finance executives, community college presidents”, from about 450 institutions (http://www.eab.com/About-Us/What-We-Do).

The report involved the input and advice from nearly 100 leaders in academic administration, including some from libraries (University Leadership Council 2011, vi – vii). The general tenor of the document is that the publishing (and therefore the library) environment is changing and that libraries need to change as well.

There is an underlying assumption in the report of widespread availability of scholarly resources as e-books, which is not borne out by reality as the results of this study demonstrate. In part this assumption of availability is based upon a mixture of hard data, such as the size and scope of the Google Books / HathiTrust project, and a more widely held set of beliefs encouraged by the extensive use of the Kindle and ubiquity of Internet access. These assumptions can be seen in statements such as:

With the rise of companies like Google and Amazon, as well as nonprofits like Wikipedia and HathiTrust users now meet most of their information needs through sources outside the library. The collections of articles, monographs and e-books made available through these organizations dwarf library collections in size and scope and content is increasingly accessed virtually through web- and cloud based distribution portals. (University Leadership Council 2011, viii)

While it is true that many scholars make frequent use of these sources, it does not acknowledge the access problems presented by copyright, licensing and the technical challenges for users. This lack of acknowledgement of these issues can be seen in the statement:

OCLC’s Constance Malpas recently studied HathiTrust’s collection in depth and found that (assuming a workable licensing model for accessing in-copyright works is established) [emphasis added] …its archive already duplicates a significant portion of even the largest research libraries and could therefore allow considerable space and storage cost savings for those who elect to withdraw volumes held electronically in HathiTrust. (University Leadership Council 2011, 22).

This paper proposes the creation of a phrase to cover this gap between what has been digitized and is available and existing academic library collections. The phrase ‘practical overlap’ is intended to describe the degree to which an academic institution can rely on electronic sources to replace its print monograph collection.
As the academic library community is well aware, the non-librarian impression of widespread availability of existing print works is highly inaccurate. The large growth in e-books has been driven by a combination new titles, popular reading, vanity publishing and public domain works. In some cases publishers are introducing back lists of titles, but these tend to be high use ‘classics’ or relatively recently published titles; often post 1995.

Much electronic publishing is tied to dedicated devices such as Kindle or Nook e-readers, and single user paradigms which are inaccessible to libraries. Although there is a large number of current and recent popular non-fiction and fiction titles available through the Kindle device and Apple Store, the licensing often does not allow library use.

**Previous research**

Several previous studies compared e-book availability with print holdings. For example, Price and McDonald (2008) compared current acquisitions of five academic libraries with catalogue data from four aggregators in 2006 and 2007, and found that about 30% of new material was available. More recently, Pomerantz (2010, 12) tested recent acquisitions from Adelphi University against the YBP acquisitions GOBI database, and found 31% of the titles received in print format were available as e-books. Link (2012) took a slightly different approach. He identified highly circulated monographs from The College of New Jersey Library, from the years 2009 and 2010, and assessed e-book availability for these titles. The finding of Link’s study was that 17% of these titles were available as e-books in that time (2012, 143).

**The case institution**

This study aims to investigate the degree of overlap between academic library print collections and e-books which are available to academic libraries. A single institution has been used to provide a typical sample built on an existing library collection.

The case institution is RMIT University (www.rmit.edu.au). The University is a large (more than 75,000 students) multi-disciplinary, doctoral level institution with three campuses in Melbourne, Australia, two campuses in Vietnam, and partner institutions in other countries. The University’s strengths are in design and technology. The former institute of RMIT gained University status in 1992, concurrent with an amalgamation of a number of smaller institutions.

The University Library has a relatively small collection of about 782,000 monographs. This figure includes approximately 137,000 e-books and 331,151 unique print monograph titles. On a per student basis, this translates to 10.71 items per student, in comparison with the Australian average of 22.6 (Council of Australian University Librarians 2011) and the Association of Research Libraries’ average of 165 items per student (2012).

As a result of the inner city location, the Library also has relatively little floor space with 0.21 square metres per student, half that of the Australian average of 0.42 square metres per student (Council of Australian University Librarians 2011). At the time of writing, proposed building alterations could also result in the temporary loss of up to 50% of the space available in the largest library which accounts for about 66% of all use across the system. This potential loss of space provided impetus for a project that measured the degree to which the physical book collection could be replaced by e-books.
Methodology

As information was being sought about replacement of existing print material with online versions, it was decided that it was necessary to compare a sample of all current print holdings against electronic sources. The approach used was similar to that taken by Pomerantz (2010), which involved manual checking of titles. There were two critical differences between this study and that undertaken by Link (2012), Pomerantz (2010) and Price and McDonald (2008). The first variation from previous studies was that it explored overlap with all existing print monograph holdings. This was necessary to compare e-holdings with the entire print monograph collection as space savings were being considered, and therefore it was necessary to include all monographs, not just current acquisitions. The second variation was that this study focused on a randomly selected sample of titles because it was not possible to check all 331,515 title records manually. Automated ISBN matching was not used as there was concern that incorrect and variant ISBNs, particularly between print and e-versions, may have provided inaccurate results.

Calculating sample size

The use of a sample required adherence to a methodologically sound randomization process. This adherence, coupled with a carefully calculated sample size produced a result with a confidence level of 95%. The sample size was calculated using the technique described by Scheaffer, Mendenhall and Ott (1996, 98).

Two variables had to be determined in order to use this technique. The first variable, the unique monograph title count of 331,515 was obtained from the University Library’s Voyager ILMS. This is referred to as “t” in the formula below. The second variable was the likely proportion of titles available as e-books. In order to obtain this figure a preliminary random sample of 100 titles was tested (using the randomization process specified below). Of this preliminary sample of 100 titles, 21 titles were available in an e-format. This variable is referred to as “p” in the formula.

In the following formula: 

\[ n = \frac{t \times p \times q}{(t-1) \times 0.000625 + p \times q} \]

Where:

\[ t = \text{number of unique print monographs in the collection} \]

\[ p = \text{the likely proportion of e-books} \]

\[ q = 1 - p \]

\[ n = \text{the number of titles to be tested in order to reach a 95\% confidence level}. \]

The calculation was applied thus:

\[ N = \frac{331,515 \times 0.21 \times 0.79}{(331,515-1) \times 0.000625 + 0.21 \times 0.79} \]

This provided a sample size of 264 titles.

Creating the sample set

In order to create a sample set of titles, the researchers then:
1. generated a list of random numbers between 1 and 10,000 using the routines available at random.org (www.random.org); and

2. searched the Voyager ILMS, using keyword search function, with each of the random numbers generated in step 1.

This search introduced an additional degree of randomization as the search was simply a string search looking for any occurrence of those digits in any searchable field, including shelf numbers, pagination statements, titles (if numerical) and bibliographic control numbers such as those generated by the National Library of Australia.

In order to avoid looking for materials which were already held in e format, as the test was intended to discover which physical items could be replaced by e-books:

3. the search was filtered by “Not e-books”; and

4. (as the catalogue as a default lists recently added material first) this list was then sorted by title, which provided date randomization.

Bibliographic details of the first print monograph entry found were then entered into an Excel spreadsheet. These details included title, author or editor, publisher, date of publication, shelf number and any other significant information such as special holdings notes. Entries for e-books, microforms and audio-visual material were ignored as these would have little or no impact on the floorspace required for shelving

**Testing**

A Master’s degree fieldwork student was assigned the task of searching each of these titles manually in several comprehensive sources. These sources were:

1. **GOBI**³: which “provides access to more than 10 million titles, including 600,000 e-books … from aggregators and publishers” (http://www.ybp.com/gobi3.html). GOBI³ is the interface used by YBP / Baker and Taylor, the University Library’s main supplier of collection materials, both print and digital. This source was chosen for its comprehensive nature as well as the applicability of content to an academic audience.

2. Amazon: which was chosen due to its pervasiveness and general popularity, as well as providing a source which might contain material not included in the YBP / Baker and Taylor database.

3. A general Google search: which was chosen to broaden the source base wider again and to pick up open access or pirated copies of material.

HathiTrust and Google Books were considered as sources, but as the large majority of that material is not available in full due to copyright law, it was decided that at the time of testing this could not be considered a viable alternative source of scholarly resources. A number of authors such as Crawford (2012), and Henry (2010, 122) point out that these sources are not viable alternatives for collection replacement primarily due copyright restrictions.
A title search was then performed on the Amazon database and on the GOBI³ database, and the presence or absence of the title as an e-book was noted on an Excel spreadsheet. In order to be considered a match, all details including edition had to match the sample bibliographic record.

**Results**

The results of the investigation were similar to those found by Link (2012), Pomerantz (2010) and Price and McDonald (2008). The study demonstrated that while there was some degree of overlap between existing print holdings and the sources examined, the rate of practical overlap was about 33%. This practical overlap rate assumes that the Library supports Amazon Kindle access via the Overdrive service, and relies on YBP for their main acquisition source. Overdrive acts as a mediator between libraries and publishers with digital rights management, lending, payment and reporting services. There has been some criticism of both Overdrive’s and publishers’ policies in the general library press. These criticisms variously have to do with lending limits, cost and availability.

**Distribution by supplier**

The following sections describe the overlap rate between each of the suppliers and the Library’s holdings, as well as overlap between suppliers.

1. GOBI³: Of the 264 titles tested, 52 (19.6%) were available as an e-book on the GOBI³ database. This is approximately the same degree of overlap between the sample title set and Amazon.

   Of the 52 titles found on GOBI³, 34 of the titles were also available on Amazon, and 17 of the GOBI³ titles (6% of the sample) were already held by the University in e-format.

2. Amazon: Of the 264 titles tested, 52 (19.7%) were available on Amazon as an e-book. Of the 52 titles found on Amazon, 34 were also available on GOBI³ (as stated above), and 13 (5% of the sample) were held by the University in electronic format, as well as being available on GOBI³. There were no Amazon only titles in the sample which were already held electronically.

Amazon presents a particular problem with the concept of availability. Currently Amazon content is available for the Kindle (and Kindle apps on devices such as i-pads), but these are generally tied to a single user account representing an individual. There has been some library licensing of Amazon content supported by Overdrive, but this is very fluid. At the time of testing, of the 52 sample titles available on Amazon, only six (2% of the sample) were available via Overdrive.

The Overdrive content still assumes that each individual reader possesses a Kindle or Kindle app and account. Due to the relatively low number of titles available under the Amazon/Kindle/Overdrive environment, the licensing, the technical complexity of users getting content onto their Kindle, and need for users to have a Kindle or Kindle app, this study regarded all Amazon content as not being part of the practical overlap.

3. Other sources: The Google search revealed that 17 titles from the sample (6%) were available from other sources - in some cases (6 titles, or 2%) were available at no charge

4. Combined availability: A total of 69 titles (26% of the sample) were available from either GOBI³ or a web-based source other than Amazon. This figure of 69 titles includes:
35 GOBI³ titles not already held;
17 GOBI³ titles already held in both print and e-book formats; and
17 titles available from other sources.

The overlap is described in the Venn diagram in Figure 1.

![Venn Diagram](image)

**Figure 1**: Availability of e-books from the sample on Amazon, GOBI³ and RMIT. (Note: all numbers refer to the number of titles available, not a percentage of titles.)

**Distribution by publication date**

Figure 2 shows the availability of e-books organised by date of publication and provider, RMIT University Library e-book holdings, and total e-book availability. The total number of titles in the sample within that range is in brackets next to the publication date.
Figure 2: The availability of e-books in the sample organised by date of publication. ABOUT HERE

Figure 3 demonstrates e-book availability as a proportion of overall sample titles for each of the periods. This chart shows that relatively recently published books, particularly those with a publication date between 2006 and 2010, are more likely to be available in a digital format than titles published prior to 2006, although those published since 2011 show a slight decline in e-book availability.

Figure 3: The ratio of e-books to print over time.
Discussion

This investigation demonstrates that it appears unlikely that more than 26% of current print holdings could be replaced with e-books. This overlap figure of 26% is similar to previous studies (Link (2012), Pomerantz (2010) and Price and McDonald (2008)), which ranged from 17% to 31% practical overlap between print books and e-books. This may be a reflection of the currency of the comparison, as the Price and McDonald study was carried out on 2006 and 2007 acquisitions, and Pomerantz on acquisitions from 2008. The Link study was performed on circulation data from 2009 and 2010. The overlap figure is also surprising as, unlike the other studies quoted, this comparison was carried out on the total collection and contained a significant proportion of material (97 titles or 37% of the sample) which was published prior to 1991.

The discovery that 17 (6.4%) print titles were already held electronically provides some encouragement for consideration of a further review of overlap between print and already purchased e-books. The Library currently has approximately 625,000 shelved items. If 6.4 % of these are already available and held as e-books, this may allow the removal (to storage or discard) of 40,000 on-shelf items. According to Courant and Nielson (2010), the average cost of holding an item on shelf for a year is US$4.26. If the Library did remove 40,000 volumes, there is a potential annual cost saving of AUD$170,000, on the assumption of $US1.00 = AUD$1.00. It must be noted that this cost includes the cost of construction as well as cleaning, electricity and a number of other costs. An alternative calculation, using a broad rule of thumb of 90 volumes per square meter of floor space, would indicate a potential savings of 444 square metres.

These same calculations of US$4.26 per volume per annum or 90 volumes per square metre could be applied to the 19% of volumes available from the principal supplier. In this case a possible savings of AUD$505,875 per annum or, in terms of floor space, 1,319 square meters could be realized. While building works at the case institution are not contemplated at this time, these figures are provided as an indication of what the consequent collection growth might save for university libraries with a policy of not discarding physical copies.

An apparently counter-intuitive aspect of the study is the relatively smaller proportion of titles available from the period 2011. Further investigation and analysis may explain why more recent titles seem to have a lower rate of e-book availability. It is possible that this reflects the fact that the Library has had an “e-preferred” policy in place for some time, and that these print titles were purchased because there was not an e-book version available at the time of ordering.

Limitations

A replication of this study using other academic library collections as a basis for the sample may yield different results as the case institution’s collection does not have a great deal of historical depth and is biased towards design and technology rather than the humanities and social sciences. Other collections with more holdings in these disciplines may produce different results, due to the possibility that e-book publications patterns vary across disciplines.

The sample size may also represent a limitation to the general applicability of these results. While the sample tested produces a 95% confidence interval, the subsets (for instance, the number of titles available from Amazon) may not be as robust.
Further research

This study gives rise to questions which would benefit from further research. Of particular value would be replication of this study in the future as it is apparent that a number of suppliers and publishers are progressively making older titles available as e-books and the practical overlap will grow. Use of an additional or alternative source such as the database of another library supplier would provide a useful contrast or confirmation of these results.

An additional and valuable investigation would involve other university libraries – either individually or in groups. This would provide a contrast which might be attributed to the difference in collection size, age and scope, or alternatively it may provide confirmation of these findings. Extending this research to other geographic regions, particularly the United Kingdom or North America would again provide confirmation or an alternative view. The bias of the collection which was tested toward design and technology would also be worth exploring as academic libraries with a greater emphasis on humanities or social sciences may produce different results.

Conclusion

This investigation was carried out to explore the degree to which an existing print monograph collection could be replicated in e-book format in order to conserve and provide additional open floorspace, particularly in relation to proposed building works. The research revealed that while some space savings would be possible, such an exercise is unlikely to release more than 26% of space which is currently in use, although due to the relatively high duplication rate of about 1.8 items per title, the space savings could be greater than 26%. These figures do not consider the multitude of associated issues, such as the advantages of e-books afforded by anywhere/anytime availability and that e-books cannot be lost, stolen, returned late or mutilated. Nor do these figures recognize the well documented preference of some readers for the physical artefact of a book or the myriad of preservation questions which surround e-books. This preference for print books may be more important in the design area in particular where images can be well presented on paper as opposed to screens.

As well as the above issues, the reality that the university would have to re-purchase each item withdrawn must be acknowledged. The magnitude of this cost alone is likely to make the exercise of replacing print-on-paper books with e-books in order to save floor space unlikely.

Acknowledgement:

The authors wish to acknowledge the contribution made by Dr Anthony Bedford, Deputy Head, School of Mathematical and Geospatial Science, RMIT University in advising on the statistical methodology adopted.

References

http://www.arl.org/stats/annualsurveys/arlstats/arlstats11.shtml


http://citesandinsights.info/v12i7a.htm


http://www.guardian.co.uk/books/2012/aug/06/amazon-kindle-ebook-sales-overtake-print


http://ccdl.libraries.claremont.edu/cdm/singleitem/collection/lea/id/161


http://library.wwu.edu/sites/library.wwu.edu/files/stratplan/redefining_academic_library.pdf