Evidence Based Library and Information Practice

A Re-examination of Online Journal Quality and Investigation of the Possible Impact of Poor Electronic Surrogate Quality on Researchers

Ken Ladd

Volume 13, Number 3, 2018

URI: https://id.erudit.org/iderudit/1102457ar
DOI: https://doi.org/10.18438/eblip29449

See table of contents

Publisher(s)

University of Alberta Library

ISSN

1715-720X (digital)

Explore this journal

Cite this article

Ladd, K. (2018). A Re-examination of Online Journal Quality and Investigation of the Possible Impact of Poor Electronic Surrogate Quality on Researchers. *Evidence Based Library and Information Practice*, *13*(3), 53–68. https://doi.org/10.18438/eblip29449

Article abstract

Objective – This study re-examines the findings of a paper (Ladd, 2010) that investigated whether evidence indicated print equivalent journal collections needed to be preserved, based on the quality of their electronic surrogates. The current study investigates whether: 1) electronic surrogate articles that failed (i.e., the print equivalent article needed to be consulted to view all the content/information) in the first study had improved in quality; and 2) there was evidence that poor-quality electronic surrogates could impact on research if the print equivalent articles did not exist.

Methods – Each of the 198 PDF documents identified in the 2010 study as failing were re-examined to assess whether any change in quality had occurred. To assess the possible impact for researchers if they needed to rely solely on poor-quality electronic journal surrogates, citation data were collected for each of the failed scholarly PDFs using Web of Science and Scopus, and usage count data were collected from Web of Science.

Results – Across the electronic journal backfiles/archives examined, there were 13.6% fewer failures of electronic surrogates for all PDF documents than in the original study, while for scholarly PDF documents (e.g., research papers) there were 13.8% fewer failures. One electronic journal archive accounted for 91.7% of the improvement for scholarly PDF documents. A second archive accounted for all the observed improvement for non-scholarly PDF documents. The study found that for the failed scholarly PDF documents from the original study, 58.7% had been cited or had Web of Science usage counts from 2010 onward.

Conclusion – The study demonstrates a continued need for retaining print equivalent journal titles for the foreseeable future, while poor-quality electronic surrogates are being replaced and digitally preserved. There are still poor-quality images, poor-quality scans of text-only articles, missing pages, and even content of PDF documents that could not be explained (e.g., incorrect text for images when compared to the print). While it is known that not all researchers will consult each of the papers that they cite, although it is best practice to do so, the extent of citations of the failed scholarly PDF documents indicate that having to rely solely on electronic surrogates could pose a problem for researchers.

Evidence Based Library and Information Practice

Research Article

A Re-examination of Online Journal Quality and Investigation of the Possible Impact of Poor Electronic Surrogate Quality on Researchers

Ken Ladd
Collection Services Librarian
University of Saskatchewan Library
Saskatoon, Saskatchewan, Canada
Email: ken.ladd@usask.ca

Received: 17 May 2018 **Accepted**: 6 July 2018

DOI: 10.18438/eblip29449

Abstract

Objective – This study re-examines the findings of a paper (Ladd, 2010) that investigated whether evidence indicated print equivalent journal collections needed to be preserved, based on the quality of their electronic surrogates. The current study investigates whether: 1) electronic surrogate articles that failed (i.e., the print equivalent article needed to be consulted to view all the content/information) in the first study had improved in quality; and 2) there was evidence that poor-quality electronic surrogates could impact on research if the print equivalent articles did not exist.

Methods – Each of the 198 PDF documents identified in the 2010 study as failing were re-examined to assess whether any change in quality had occurred. To assess the possible impact for researchers if they needed to rely solely on poor-quality electronic journal surrogates, citation data were collected for each of the failed scholarly PDFs using Web of Science and Scopus, and usage count data were collected from Web of Science.

Results – Across the electronic journal backfiles/archives examined, there were 13.6% fewer failures of electronic surrogates for all PDF documents than in the original study, while for

scholarly PDF documents (e.g., research papers) there were 13.8% fewer failures. One electronic journal archive accounted for 91.7% of the improvement for scholarly PDF documents. A second archive accounted for all the observed improvement for non-scholarly PDF documents. The study found that for the failed scholarly PDF documents from the original study, 58.7% had been cited or had Web of Science usage counts from 2010 onward.

Conclusion – The study demonstrates a continued need for retaining print equivalent journal titles for the foreseeable future, while poor-quality electronic surrogates are being replaced and digitally preserved. There are still poor-quality images, poor-quality scans of text-only articles, missing pages, and even content of PDF documents that could not be explained (e.g., incorrect text for images when compared to the print). While it is known that not all researchers will consult each of the papers that they cite, although it is best practice to do so, the extent of citations of the failed scholarly PDF documents indicate that having to rely solely on electronic surrogates could pose a problem for researchers.

Introduction

There continues to be increased demand for user space within academic libraries. In recognition of these needs and with the availability of electronic journal backfiles of content held in print by libraries, there is opportunity to repurpose prime library space once occupied by print journal collections. At the same time, preservation is still recognized as a fundamental role and responsibility of research libraries (ARL, 2007). With the goal of preserving information for future generations coupled with the desire to remove print collections from prime library space, this is often accomplished by the relocation of print materials into storage facilities, disposal of titles through participation in collaborative print archive initiatives, or the disposal of print journals where an electronic surrogate exists.

The strategy of removing print equivalent journals where an electronic surrogate exists is complicated by known quality issues with electronic surrogates (Bracke & Martin, 2005; Chen, 2005; Erdman, 2006; Hawkins & Shadle, 2004; Henebry, Safely, & George, 2002; Joseph, 2006, 2012, 2014; Kalyan, 2002; Keller, 2005; Ladd, 2010; Martellini, 2000; McCann & Ravas, 2010; Robinson, 2010; Sprague & Chambers, 2000; Thohira, Chambers, & Sprague, 2010; Weessies, 2012), where there can be missing content (volume issues or pages), poor-quality images, and illegible text from poor-quality scans. Ladd (2010) concluded that the re-digitization of failed PDF content using high-resolution technology along with good quality control practices would eliminate many of the observed failures. Given the number of studies reporting quality issues with electronic surrogates, which can be corrected by re-digitization, would publishers attempt to address this significant issue? This is important as it affects users of e-journal backfiles and libraries considering the removal of print equivalent materials from their collections.

Because it was known that there were quality issues associated with electronic journal backfiles, the author believed that over a seven-year period there had been sufficient time for publishers to address some of these issues. It was felt that revisiting the original study now could assist in the development or revision of recommendations for the preservation period of print equivalent titles.

In 2006, Elsevier began to replace poor-quality images on a case-by-case basis, which developed into an extensive initiative that resulted in hundreds of thousands of pages being

rescanned (van Gijlswijk & Clark, 2010). This raised two key questions:

- What impact has Elsevier's initiative had on the overall quality of their electronic journal backfiles?
- Have other publishers attempted to address the quality of their electronic journal backfiles and to what degree?

These questions are important, as the extent to which the quality of electronic surrogates have been improved could affect the need to preserve print equivalent titles.

Joseph (2012) followed up an earlier study of Elsevier's Earth and Planetary Sciences archive to investigate the impact of Elsevier's rescanning project. The study was, however, of one disciplinary journal archive of one publisher. The current study was designed to investigate the journal archives of multiple publishers/vendors by re-examining the results of Ladd's 2010 study. In that study, Ladd chose seven electronic journal backfiles acquired by the University of Saskatchewan that covered a breadth of subjects. Journal titles were randomly selected from each backfile and from these titles, volumes and then issues were randomly selected. Complete issues were then examined. A total of 2,633 PDF documents were examined and then compared with their print equivalents.

As noted above, the quality of electronic journal backfiles can potentially affect researchers and scholars when they attempt to access PDF documents with poor-quality images, illegible text, or missing pages. The author wanted to investigate the level of potential impact if researchers could only rely on electronic journal archives. As a proxy measure of the potential impact, the current study uses citations to scholarly articles that Ladd identified in 2010 as being of poor quality and were found to still be of poor quality in 2017.

Literature Review

Numerous researchers have investigated the differences between electronic surrogates and their print equivalents (Bracke & Martin, 2005; Campbell, 2003; Chen, 2005; Chrzastowski, 2003; Erdman, 2006; Hawkins & Shadle, 2004; Henebry et al, Safely, & George, 2002; Joseph, 2006, 2012, 2014; Kalyan, 2002; Keller, 2005; Ladd, 2010; Martellini, 2000; McCann & Ravas, 2010; Robinson, 2010; Sprague & Chambers, 2000; Thohira et al, 2010; Weessies, 2012). These studies were most often conducted to determine if the electronic surrogates allowed libraries to cancel or withdraw print equivalent titles from their libraries. The studies often focused on a specific factor such as a discipline, missing content, vendor, or electronic journal backfiles or aggregators.

Researchers have often found one or more of the following quality issues associated with the scanned electronic surrogates:

- images and figures (Bracke & Martin, 2005; Chen, 2005; Erdman, 2006; Henerby et al., 2002; Joseph 2006, 2012, 2014; Keller, 2005; Ladd, 2010; McCann & Ravas, 2010; Robinson, 2010; Sprague & Chambers, 2000; Thohira et al., 2010),
- illegible text and formulas (Keller, 2005; Ladd, 2010; Sprague & Chambers, 2000; Thohira et al., 2010),
- missing content—figures, tables, missing pages, articles or issues (Bracke & Martin, 2005; Chen, 2005; Henebry et al., 2002; Joseph, 2006; Keller, 2005; Ladd, 2010; Sprague & Chambers, 2000; Thohira et al., 2010).

Campbell (2003) found no substantial content missing for the titles reviewed. Chrzastowski (2003) noted that while quality was still a concern, over a two-year period there had been only one problem for the chemistry and chemistry-related e-journals at University of

Illinois at Urbana-Champaign, and that the vendor had quickly addressed the problem.

Ladd (2010) noted that many of the quality related issues observed in the study could be resolved if the existing electronic surrogates were replaced with scans using higher-resolution scanning technology and better quality control. As noted previously, in 2006 Elsevier began replacing poor-quality images on a case-by-case basis. This ultimately led to a large-scale initiative that saw hundreds of thousands of pages with poor-quality images being rescanned (van Gijlswijk & Clark, 2010).

There has been one study that re-examined the observed problems with the quality of electronic surrogate journals. Joseph (2006) conducted a study of 35 titles in Elsevier's Earth and Planetary Sciences archive and found that 73.6% of the volume issues had at least one figure that was of poor quality. In a follow-up study to investigate the impact of Elsevier's rescanning project, the number of issues with poor-quality images was extrapolated to have been reduced to 21.9% (Joseph, 2012). The study was, however, of one electronic journal backfile from one publisher, in a disciplinary area whose papers often contain images. By contrast, the current study is multi-disciplinary, re-examining seven different electronic journal archives, with a number of different publishers to determine whether there has been an improvement in the quality of the electronic surrogates. In addition, by examining the potential impact on researchers if they needed to rely solely on poor-quality electronic surrogates, this study fills an important need since there have been no other studies of this nature.

Aim

This study investigated whether there continues to be evidence print equivalent serials need to be preserved for the short to medium term because of poor-quality electronic surrogates, as concluded in a previous study (Ladd, 2010). The central questions were:

- Have the PDF documents that failed in the 2010 study subsequently improved in quality?
- Were there differences in the improvement of quality between electronic surrogate archives?

A second objective of this study was to examine whether there was evidence that having to rely solely on electronic surrogates could potentially impact researchers. To examine this issue, the study asked, for PDF documents observed to have failed in the 2010 study and found to still fail in 2017:

- What citations have occurred from 2010 onward?
- Is there evidence of their usage?

Methods

The original 2010 study examined PDF documents from seven electronic journal backfiles (Appendix) from a number of vendors with a breadth of subject coverage (humanities, social sciences, science, technology, and medicine). In that study, a PDF document from an electronic surrogate was assessed as failing any time the print equivalent needed to be consulted in order to gain access to all of the item's information. In the current study, each of the PDF documents from the original study that were classified as failing served as the study sample.

In the fall of 2017, each of the 198 PDF documents that failed in the original study was downloaded from the publisher's backfile and re-examined to determine if it still was classified as failing, using the original definition for a failure. Data were collected for each collection archive and journal title examining:

- the number and percentage of the 174 PDF previously failed documents with scholarly content, which had failed

again. Scholarly content included research papers, case studies, review articles, short communications, technical notes, and errata.

- the number and percentage of the 24 PDF previously failed documents with other content, which had failed again. Other content included book reviews, announcements, letters to the editor, meeting programs, front and back matter, and obituaries.

These data were compared to the 2010 data to determine whether there had been an improvement in the quality of the electronic surrogates and for which electronic journal collection backfiles.

The second part of the study examined researchers' consultation of the 150 scholarly PDF documents that were identified as still failing in the current study. These papers were published between 1938 and 1999. Two proxies for consultation of these articles were used: 1) citation of the failed PDF documents from 2010 onward using citation data from Web of Science and Scopus, and 2) the usage count feature of the Web of Science, which records the number of times that the full-text of a record has been accessed or where a record has been saved by any Web of Science user in the last 180 days or since February 1, 2013.

Results

Ladd (2010) found that there were 198 PDF documents that were assessed as failing—174 were scholarly and 24 consisted of other content such as book reviews and announcements. When each of these PDF documents was examined for the current study, some improvement in the quality of the electronic surrogates was observed. Table 1 provides data on the frequency of failures for PDF documents (all PDFs, scholarly PDFs, and other PDFs) for the original study and the current study by electronic journal archive collection.

Table 1
Failed Electronic Surrogates (All, Scholarly, and Other PDF Documents), 2017 Compared to 2010

Collection	All PDFs		Scholarly PDFs		Other PDFs	
	2010	2017	2010	2017	2010	2017
Elsevier Science Direct Backfile - Medicine and Dentistry	67	45	62	40	5	5
Elsevier Science Direct Backfile - Social Science	25	24	20	19	5	5
JSTOR Arts and Science I	5	2	1	1	4	1
Oxford University Press Digital Archive	19	19	12	12	7	7
Springer Link Archives (Mathematics)	3	2	2	1	1	1
Wiley - Humanities and Social Sciences	32	32	32	32	0	0
Wiley - Science, Technology and Medicine	47	47	45	45	2	2
Total	198	171	174	150	24	21
PDFs - Improved quality		**13.6%**		**13.8%**		**12.5%**

For scholarly PDF documents, 13.8% (24) were no longer found to have failed. The results indicate, however, that all but two of the 24 documents that now passed were from a single archive, Elsevier ScienceDirect Backfile - Medicine and Dentistry (a 35.5% improvement in quality). The Elsevier ScienceDirect Backfile - Social Science collection and Springer Link Archives (Mathematics) each had a single document that no longer failed.

For the other PDF documents, 12.5% (3) were no longer found to have failed, all from the JSTOR Arts and Science I archive. This represented a 75% increase in quality for this archive.

The original study noted that scholarly PDFs failed for a variety of reasons: quality of graphs, maps or drawings; illegible text/numbers in a table or article; missing or incorrect images or content; and quality of the image. Figure 1 illustrates the frequency of scholarly PDF documents failing in the 2010 study and the current study for the Elsevier ScienceDirect Backfile - Medicine and Dentistry archival collection, by type of failure: quality control (pages missing or incorrect images), other (illegible text, tables, drawings, or graphs), or image (e.g., x-rays, scintigraphs, photographs, and others).

The study found that each of the PDFs that were now observed to pass had failed originally because of poor-quality images. This represents a 52.4% decrease in the number of failures because of image quality. For two of the PDF documents that still failed, there had been multiple images in each that were of poor quality in the original study, but for the current study all but one of the images in each PDF were now of good quality.

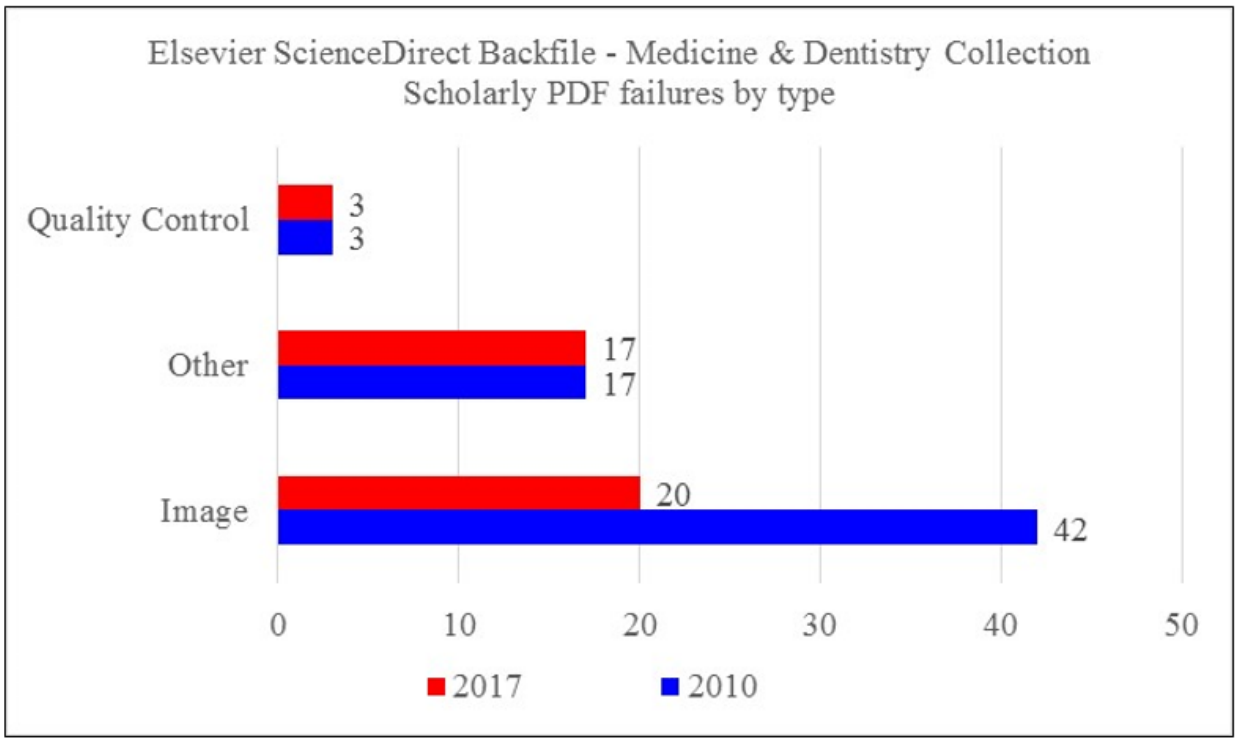

Figure 1
Comparison of failures by type for scholarly PDFs between the two studies for the Elsevier ScienceDirect Backfile - Medicine and Dentistry collection.

Table 2
Failed Scholarly PDFs Cited and Total Number of Citations in Web of Science and Scopus from 2010 Onward

		WOS				Scopus			
Collection	**Failures**	**Cited**	**%**	**Citations**	**Citations / Cited Article**	**Cited**	**%**	**Citations**	**Citations / Cited Article**
Elsevier Science Direct - Medicine and Dentistry	40	15	37.5%	45	3.0	17	42.5%	54	3.2
Elsevier Science Direct - Social Science	19	16	84.2%	156	9.8	18	94.7%	186	10.3
JSTOR Arts and Science I	1	1	100.0%	1	1.0	0	0.0%	0	0.0
Oxford University Press Digital Archive	12	5	41.7%	13	2.6	5	41.7%	13	2.6
Springer Link Archives - Mathematics	1	1	100.0%	3	3.0	1	100.0%	3	3.0
Wiley Blackwell Backfiles - Humanities and Social Sciences	32	9	28.1%	669	74.3	13	40.6%	845	65.0
Wiley Blackwell Backfiles - Science, Technology and Medicine	45	23	51.1%	162	7.0	24	53.3%	176	7.3
TOTAL	**150**	**70**	**46.7%**	**1049**	**15.0**	**78**	**52.0%**	**1277**	**16.4**

Table 3
Failed Scholarly PDFs Cited and Total Number of Citations Unique between Web of Science and Scopus from 2010 Onward

Collection	Failures	Cited	%	Citations	Citations / Cited Article
Elsevier Science Direct - Medicine and Dentistry	40	17	42.5%	64	3.8
Elsevier Science Direct - Social Science	19	18	94.7%	202	11.2
JSTOR Arts and Science I	1	1	100.0%	1	1.0
Oxford University Press Digital Archive	12	5	41.7%	16	3.2
Springer Link Archives - Mathematics	1	1	100.0%	3	3.0
Wiley Blackwell Backfiles - Humanities and Social Sciences	32	13	40.6%	981	75.5
Wiley Blackwell Backfiles - Science, Technology and Medicine	45	26	57.8%	182	7.0
TOTAL	**150**	**81**	**54.0%**	**1449**	**17.9**

The current study examined the potential impact for researchers if they could consult only the poor-quality electronic surrogates. One proxy for possible impact is the citations from 2010 onward to the scholarly PDF documents that were observed to still have failed in the current study. For Web of Science, Scopus, and unique (between the two databases), Tables 2 and 3 present the number of failed PDFs that had been cited from 2010 onward, the total number of citation counts for all PDFs, and percentage of failed articles cited for each electronic journal archive.

A total of 81 (54.0%) of the failed PDFs had been cited from 2010 onward, the year the first study was published. There were 1,449 unique citations for these 81 papers, however one paper accounted for 654 of the citations. The remaining 80 papers had 795 citations or an average of 9.9 citations each. For the five archival collections with more than 10 failed scholarly PDFs, the percent cited ranged from 40.6% to 94.7%. Regardless of the disciplinary area, a significant number of the failed PDFs were cited.

The study examined the Web of Science usage count feature as a second proxy for the possible impact of researchers having to consult only poor-quality electronic surrogates. The Web of Science database (Web of Science Core Collection Help, 2018) defines usage as any Web of Science user either "…clicking links to the full-length article at the publisher's website (via direct link or Open-URL) or by saving the article for use in a bibliographic management tool (via direct export or in a format to be imported later)." Table 4 presents the Web of Science

Table 4
Failed Scholarly PDFs with Web of Science Usage Count by Collection

		Failed PDFs with WOS usage data		
Collection	**Failures**	**Not cited**	**Total**	**Percent**
Elsevier ScienceDirect – Medicine and Dentistry	40	4	11	27.5%
Elsevier ScienceDirect – Social Science	19	2	12	63.2%
JSTOR Arts and Science I	1	0	0	0.0%
Oxford University Press Digital Archive	12	0	1	8.3%
Springer Link Archives - Mathematics	1	0	0	0.0%
Wiley Blackwell Backfiles - Humanities and Social Sciences	32	0	0	0.0%
Wiley Blackwell Backfiles - Science, Technology and Medicine	45	1	12	26.7%
TOTAL	**150**	**7**	**36**	**24.0%**

usage count data for the scholarly PDFs that were found to still have failed in this study: number not cited in Web of Science or Scopus, total number with usage data, and percentage of the total failures.

The study found that 36 (24.0%) of the 150 failed scholarly PDFs had Web of Science usage data associated with them. Of these 36, seven had no citations in Web of Science or Scopus. Using the two proxies for possible impact of consulting only poor-quality electronic surrogates, there were 88 (58.7%) failed scholarly PDFs that had either citations or Web of Science usage data from 2010 onward.

Discussion

The current study found that only one electronic journal archive collection, Elsevier ScienceDirect Backfile - Medicine and Dentistry, had improved significantly in quality since the original 2010 study. In that collection, more than one-third (35.5%) of the failed scholarly PDFs were now observed to not fail. Of the remaining electronic archival collections, only two had any improved scholarly PDFs: Elsevier ScienceDirect Backfile - Social Science collection and Springer Link Archives (Mathematics) each having a single scholarly PDF that no longer failed. Figure 1 shows that all the scholarly PDFs that were observed to no longer fail for Elsevier

ScienceDirect Backfile - Medicine and Dentistry failed originally because of poor-quality images.

The Elsevier rescanning project focused on pre-1995 journals, using an algorithm to identify automatically poor-quality scanned images (van Gijlswijk & Clark, 2010). The initiative analyzed 19 million pages and resulted in the rescanning of 600,000 pages of poor-quality images. All of the Elsevier ScienceDirect Backfile - Medicine and Dentistry papers in this study were pre-1995 and it would appear that this archive's 52.4% reduction in failures because of poor-quality images is linked to the Elsevier rescanning project. Compared to the current study, Joseph (2012) found greater improved quality resulting from the Elsevier rescanning initiative, likely the result of different study methodology, examining a different Elsevier electronic journal backfile, and the timing of the original and re-examination studies. Joseph's studies were done prior to and after the Elsevier initiative. Ladd's (2010) original study was done while the Elsevier initiative was moving toward completion. The results of both studies, however, demonstrate that good scanning technology coupled with good quality control practices would help to eliminate the majority of observed poor-quality scans.

Although the strategy employed by Elsevier was successful in addressing many of the poor-quality images, there are still poor-quality images and line drawings, along with other issues found by Joseph (2012) and the current study. An excellent example of problems that still exist was found in a single paper from Elsevier Science Direct Backfile – Medicine and Dentistry. When compared to the print equivalent paper, this scholarly PDF was found to be missing six of 12 plates of images (radiographs, micrographs or photograph), each with two figures per plate. For the six plates that were included in the e-surrogate, four plates or eight figures had the incorrect image associated with the description below the figure. For example, Plate XVIII had the descriptions for Figure 8 and 9, but had the images for Figure 12 and 13 of the print paper. Two of the plates had images for the figures that were upside down, and for one of these plates, the incorrect figure appeared above the description. To verify that the print copy in hand was not the aberration, several interlibrary loan copies were acquired from other academic institutions, which were determined to be identical in content to the print copy in hand.

There are a number of approaches that can be taken to address the problem of poor-quality scans, but there are significant challenges and costs associated with each. Rescanning whole issues of journals is a very time-consuming and costly approach, as is trying to find and replace poor-quality scanned pages, which are often scattered and in a minority amongst the acceptable quality scans (Joseph, 2012). Elsevier's algorithmic strategy to help address the cost associated with identifying digitized articles with poor-quality images required running the algorithm on two dedicated servers, 24 hours a day, 7 days a week for almost two years (van Gijlswijk & Clark, 2010).

A more cost-effective approach would be to crowd-source the identification of poor-quality scans that should be replaced. Researchers, readers, librarians, and others during the course of their activities could identify and report poor-quality scans to publishers as they are found, who can then replace the poor scans. This would greatly reduce the cost of identifying poor-quality scans of all types. The cost to rescan these pages would remain, however. Joseph (2012) cautioned that even after massive efforts, such as Elsevier's project to address the issue, problems with poor-quality images continue, which should be taken into consideration when making decisions to store or discard print equivalent titles. The implication is that archiving of print journal runs will be needed for the foreseeable future.

Since there continues to be a need for the preservation of print for the foreseeable future, a collaborative approach would logically be the

most cost effective, by sharing the cost of archiving amongst many institutions. For this reason, collaborative print journal storage initiatives have existed for many years around the world, allowing participating institutions to remove these titles from prime library space. However, depending on the collaborative strategy being used, there are still potential issues, even while there are undeniable benefits. The collaborative approach is excellent for sharing costs, but unless a page-by-page review is conducted of the items being archived, along with the archiving of best copy, there is a risk of archiving a damaged copy. This could prevent the rescanning of specific journal articles, should it be needed, depending on where the damage exists.

As part of the current study, the benefits of a collaborative approach were demonstrated while consulting the print equivalent volumes held at the University of Saskatchewan to compare electronic backfile and print equivalent content. It was discovered that since the 2010 study four titles had been removed from the University of Saskatchewan collection. Each of these titles were part of the Council of Prairie and Pacific University Libraries Shared Print Archive Network initiative. While the titles were no longer at the University of Saskatchewan, they were held at partner institutions. The volume issues were able to be examined at the archive partner institutions. In one case, however, the title was not found at the initial archive partner consulted, but was available at the second archive holder. This may have been because the title was in the process of being transferred to the institution's storage facility, but this example demonstrates the importance of having multiple archived copies.

While this study and others have shown that there are issues with the quality of electronic surrogates of print journal articles, there is a question of the extent of the impact to researchers if they had to rely solely on poor-quality electronic surrogates. In the current study, the author used two proxies to estimate the possible impact of poor-quality electronic surrogates. The first examined the citations to electronic surrogates of articles that were found in this study to fail. With 54% of the electronic surrogates having citations since the 2010 original study, it is apparent that many of the papers are still being actively consulted and referenced. On average, there were 9.9 citations per paper when the one paper with over 600 citations is not included in calculating the average.

The second proxy for impact was the Web of Science usage count feature. There were 36 or 24% of the failed PDFs with Web of Science usage. Of these, seven also did not have citations from 2010 onward, bringing the total to 88 papers or 58.7% of the failed PDFs with citations or Web of Science usage data. The author found, however, that the Web of Science usage data had some issues with reliability. The original data were collected in early 2016 and in preparation for writing this paper, were refreshed in early 2017. The author was surprised to note some decreases in the Web of Science usage data gathered since 2013. It was logical that usage would only increase over time. Yet for 26 papers, this figure actually decreased. Clarivate was contacted and asked why this might be the case. Clarivate responded that in April 2016, they had identified a new type of bot activity and they had adjusted their algorithms to account for the elevated usage counts (personal communication, April 4, 2017). The result was a usage count reduction to zero for 17 of the 26 affected papers.

The proxy measures for impact, particularly citations, demonstrate that researchers use the failed papers actively. The degree of impact if authors had to rely solely on poor-quality electronic surrogates will be dependent on whether the researcher needs to consult the image, text, or content in the paper that is of poor quality or missing. Regardless, with 58.7% of the failed papers being cited or having Web of Science usage data from 2010 onward, the current study indicates that relying solely on

electronic surrogates has a potentially significant impact on researchers when the electronic surrogate is of sufficiently poor quality to require consulting the original print version.

Conclusion

This study was undertaken to determine whether evidence of electronic surrogate quality continued to support the need to preserve print equivalent journals collections. Evidence was sought by re-examining PDF documents that had been classified as failing in a previous study (Ladd, 2010) to determine if their quality had improved. The study also examined whether there was evidence of potential impact on researchers if they relied only on poor-quality electronic surrogates. An indication of the extent of the potential impact was first examined by tallying the citations to scholarly PDF documents that were observed to continue to fail in the current study, and second by recording their Web of Science usage counts.

The data demonstrate clearly that there continues to be an issue with the quality of PDFs held in electronic journal backfiles. Almost all of the scholarly PDFs that no longer failed came from a single electronic journal archive (Elsevier Science Direct Backfile – Medicine and Dentistry), following a massive project conducted by the publisher to identify and replace poor-quality images. Despite Elsevier's initiative being successful in addressing many of the poor-quality images, this study still observed numerous poor-quality images and other problems in their backfiles.

An alternate approach to the one used by Elsevier, and likely more cost effective, may be a collaborative approach among vendors, libraries, and users to identify poor-quality scholarly PDFs and replace them with high-quality, high-resolution PDFs. Joseph (2012) suggested that Elsevier should at a minimum provide a form on their website to allow readers and librarians to report quality issues and incorporate addressing the reported problems into their workflows. A crowd-sourcing approach would help address the costs associated with reviewing and identifying scanned PDFs with poor-quality images, graphs, line drawings, and text. In addition, this approach would identify where poor quality control has resulted in content missing or being incorrect. While not a comprehensive strategy to address all of the quality issues with scanned journal PDFs, it would identify problems as the publications are being used, an indicator of potential future use.

Because of the cost, time, and money to address this significant problem of poor-quality scanned journal PDFs, it can be concluded that it will persist for the foreseeable future and thereby require the preservation of print serials. Thus, it would be desirable to have a comprehensive strategy that ensures that there are complete preserved copies available. One way to ensure this objective would be to use page-by-page verification for each preserved journal volume and issue. Due to the costs in time and money, this strategy is not likely to be used extensively, but if implemented would be best achieved through a collaborative approach to share the resource implications. As a less expensive alternative, redundancy for any given title among different preservation initiatives would logically compensate for less rigorous content verification. This strategy, however, does carry its own costs since it would require a greater number of copies to be preserved.

Collaborative print journal storage initiatives have existed for numerous years. This study and others indicate that there will be an ongoing need for print equivalent storage for the foreseeable future. While there have been papers written about individual initiatives and about initiatives in general, it would be of value to study at least a cohort of these initiatives to have data, for example, on their extent, retention period commitments, and validation method employed. This will shed light on whether the initiatives collectively are achieving a level of print preservation for these resources that will

help to ensure that quality print journals are available, to allow for consultation or rescanning should the need arise.

References

Association of Research Libraries. (2007). *Research libraries' enduring responsibility for preservation*. Retrieved April 16, 2018 from http://www.arl.org/bm%7Edoc/preservation_responsibility_24july07.pdf

Bracke, M. S., & Martin, J. (2005). Developing criteria for the withdrawal of print content available online. *Collection Building*, *24*(2), 61–64. http://doi.org/10.1108/01604950510592670

Campbell, S. (2003). Print to electronic journal conversion: Criteria for maintaining duplicate print journals. *Feliciter*, *49*(6), 295–297.

Chen, X. (2005). Figures and tables omitted from online periodical articles: A comparison of vendors and information missing from full-text databases. *Internet Reference Services Quarterly*, *10*(2), 75–88. http://doi.org/10.1300/J136v10n02_07

Chrzastowski, T. E. (2003). Making the transition from print to electronic serial collections: A new model for academic chemistry libraries? *Journal of the American Society for Information Science and Technology*, *54*(12), 1141–1148. http://doi.org/10.1002/asi.10318

Erdman, J. M. (2006). Image quality in electronic journals: A case study of Elsevier geology titles. *Library Collections, Acquisitions, and Technical Services*, *30*(3–4), 169–178. http://doi.org/10.1016/j.lcats.2006.08.002

Hawkins, L., & Shadle, S. (2004). Electronic journal forum: Reflections on wrapping paper: Random thoughts on AACR2 and electronic serials. *Serials Review*, *30*(1), 51–55. http://doi.org/10.1080/00987913.2004.10764877

Henebry, C., Safley, E., & George, S. E. (2002). Before you cancel the paper, beware: All electronic journals in 2001 are NOT created equal. *The Serials Librarian*, *42*(3 & 4), 267–273. http://doi.org/10.1300/J123v42n03_17

Joseph, L. (2006). Image quality in electronic journals: A case study of Elsevier geology titles. *Library Collections, Acquisitions, and Technical Services, 30*(304), 169-178. http://doi.org/10.1016/j.lcats.2006.12.002

Joseph, L. (2012). Improving the quality of online journals: Follow-up study of Elsevier's backfiles image rescanning project. *Library Collections, Acquisitions, and Technical Services, 36*(1), 18-23. http://doi.org/10.1016/j.lcats.2011.08.001

Joseph, L. E. (2014). Image quality in University of Illinois digital geology dissertations from ProQuest. *Issues in Science and Technology Librarianship*, (77). http://doi.org/10.5062/F4Z31WM1

Kalyan, S. (2002). Non-renewal of print journal subscriptions that duplicate titles in selected electronic databases: A case study. *Library Collections, Acquisitions, and Technical Services*, *26*(4), 409–421. http://doi.org/10.1016/S1464-9055(02)00287-7

Keller, A. (2005). The race to digitize: Are we forfeiting quality? *Serials, 18*(3), 211–217. http://doi.org/10.1629/18211

Ladd, K. F. (2010). An examination of the rate and content equivalency of electronic surrogates and the implications for print equivalent preservation. *Evidence Based Library and Information Practice*, *5*(4), 7-20. http://doi.org/10.18438/B83P6V

Martellini, E. (2000). Physics journals and their electronic version: A comparison. *High Energy Physics Libraries Webzine*, (2). Retrieved from http://webzine.web.cern.ch/webzine/index.html

McCann, S., & Ravas, T. (2010). Impact of image quality in online art history journals: A user study. *Art Documentation: Journal of the Art Libraries Society of North America*, *29*(1), 41–48. http://doi.org/10.1086/adx.29.1.27949538

Robinson, A. (2010). University of Kansas print and electronic journal comparison study. *Art Documentation: Journal of the Art Libraries Society of North America*, *29*(1), 37–40. http://doi.org/10.1086/adx.29.1.27949537

Sprague, N., & Chambers, M. B. (2000). Full text databases and the journal cancellation process: A case study. *Serials Review*, *26*(3), 19–31. http://doi.org/10.1080/00987913.2000.10764597

Thohira, M., Chambers, M. B., & Sprague, N. (2010). Full-text databases: A case study revisited a decade later. *Serials Review*, *36*(3), 152–160. http://dx.doi.org/10.1080/00987913.2010.10765304

van Gijlswijk, E., & Clark, B. (2010). ScienceDirect upgrades 600,000 backfiles pages. *Elsevier Library Connect*, *8*(1), 4. Retrieved from http://libraryconnect.elsevier.com/sites/default/files/lcn0801.pdf

Web of Science Core Collection Help. (2018). Retrieved March 23, 2018, from http://images.webofknowledge.com//WOKRS529AR7/help/WOS/hp_usage_score.html

Weessies, K. W. (2012). Local history maps in full text resources. *Journal of Map and Geography Libraries*, *8*(3), 230–241. http://doi.org/10.1080/15420353.2012.700300

Appendix
Titles Compared in Each Collection

Elsevier Science Direct Backfile Medicine and Dentistry
- ✓ *American Journal of Orthodontics*
- ✓ *Biochemical Medicine and Metabolic Biology*
- ✓ *British Journal of Tuberculosis and Diseases of the Chest*
- ✓ *International Journal of Nuclear Medicine and Biology*
- ✓ *Prostaglandins, Leukotrienes, and Medicine*

Elsevier Science Direct Backfile Social Sciences
- ✓ *Government Publications Review*
- ✓ *Journal of Behavioral Economics*
- ✓ *Social Science & Medicine. Part B, Medical Anthropology*
- ✓ *Studies in Comparative Communism*
- ✓ *Transportation Research. Part A, General*

JSTOR Arts and Sciences 1
- ✓ *American Journal of Mathematics*
- ✓ *Journal of Health and Human Behavior*
- ✓ *Journal of the History of Ideas*
- ✓ *Reviews in American History*
- ✓ *Speculum*

Oxford University Press Journals Digital Archive
- ✓ *Occupational Medicine*
- ✓ *Parliamentary Affairs*
- ✓ *Past & Present*
- ✓ *Rheumatology*
- ✓ *The Year's Work in Clinical and Cultural Theory*

Springer Link Archive (Mathematics Archive)
- ✓ *Computational Optimization and Applications*
- ✓ *Constraints*
- ✓ *Journal of cryptology*
- ✓ *Journal of nonlinear science*
- ✓ *K-Theory*

Wiley Blackwell Backfiles - Humanities and Social Sciences (acquired as Wiley Interscience (Synergy Blackwell) – Humanities and Social Sciences backfile)
- ✓ *Papers in Regional Science*
- ✓ *Social Policy and Administration*
- ✓ *Journal of Philosophy of Education*
- ✓ *Review of Policy Research*

Wiley Blackwell Backfiles - Science, Technology and Medicine (acquired as Wiley Interscience (Synergy Blackwell) – Science, Technology and Medicine backfile)
- ✓ *European Journal of Clinical Investigation*

- ✓ *International Journal of Experimental Pathology*
- ✓ *Journal of Human Nutrition and Dietetics*
- ✓ *Journal of Oral Pathology and Medicine*
- ✓ *Sedimentology*