2016

The persistence of open access electronic journals

Elizabeth A. Lightfoot
Florida International University, elightfo@fiu.edu

Recommended Citation

The Persistence of Open Access Electronic Journals

Author Details:

Elizabeth A. Lightfoot

University Libraries; Florida International University; Miami; Florida; USA

elightfo@fiu.edu

Biographical Details:

Elizabeth Lightfoot is the electronic resources librarian and serials coordinator at Florida International University. She has a background in serials acquisitions, with primary interest in the management and sustainability of serials and e-resources.

Structured Abstract:

Purpose – Open access (OA) electronic journals have been identified as potentially at risk of loss without more coordinated preservation efforts. The purpose of this paper is to test the current availability of OA electronic journals indexed in the Directory of Open Access Journals (DOAJ).

Design/methodology/approach – Using publicly available journal metadata downloaded from DOAJ, individual journal URLs were tested for validity and accessibility using a Microsoft Excel Visual Basic for Applications macro.

Findings – Initial results showed 69.51% of the URLs tested returned a successful HTTP status code. The remainder of the URLs returned codes that indicated redirection or errors.

Originality/Value – Unlike past studies of link decay, this is not limited to cited references or a specific discipline. This study utilizes the full DOAJ metadata to analyze the persistence of OA electronic journals.

Paper type: Research paper

Introduction

Open access publishing is moving at an increasingly quick pace, and many new born-digital journals are being added to the scholarly record. From 2003 to 2015, almost 10,000 journals were added to the Directory of Open Access Journals (Olijhoek *et al*., 2015). As librarians struggle to track the perpetual access rights of licensed content, much less time and effort has been spent on the tending of open access (OA) electronic journals. Librarians, however, continue to identify the OA electronic journals of smaller publishers and associations as content that stands the greatest risk of loss if pro-active steps towards archiving and preservation are not taken.

Literature review

Since the advent of the electronic journal, librarians and publishers have been busy charting new territory, and the conversation has primarily been focused on business models, staffing, and workflows. The increase in the availability of electronic resources, and the ephemeral nature of electronic information itself, demanded information professionals move quickly to meet the changing needs and preferences of users, as well as master the tasks related to the administration of new formats and growing digital library collections (Leibowitz, 2002; McMillan *et al.*, 1991).

For several decades, librarians have been just a few of the voices among many to see the beauty and excitement, as well as complications and the potential for loss, in the electronic medium. In 1987, Jay David Bolter imagined "the electronic reader [moving]…easily across thousands of different books or articles as he moves from one page to the next" (p. 15), but he also posited the instability of text in a computer. In 1991, librarians at Virginia Polytechnic Institute shared prescient hopes for the success of electronic journals, predicting the rise of the new medium and its imminent costs, and they wondered if library "procedures [would] handle what is REALLY coming in electronic publishing" (McMillan *et al*., p. 81). In 1995, an article in Scientific American considered the application of traditional archival methods to electronic files, and asserted that "documents may be irretrievably lost to future generations if we do not take steps to preserve them now" (Rothenberg, p. 42). Fifteen years later, Judy Luther (2010) moderated a NASIG Conference panel that discussed the need for "a deliberate act" (p. 74) to ensure future access to electronic journals. Cornell University Librarian Anne Kenney presented very similar concerns at the 2015 NASIG Conference, citing current preservation efforts as "not adequate" (Kenney and Wesley, 2016, p. 73).

Questions about long-term access and responsibility in this brave new world arose early and linger today (Beh and Smith, 2012; Duranceau, 1998; Keyhani, 1998; Morris, 2000; Szydlowski, 2010). The concern of losing parts of the scholarly record or, to a greater extent, human and cultural

consciousness is not new, nor is it unique to the electronic medium. Physical information and collections are threatened by war, acts of nature, time, or accidents (Arnold, 1997; Barrow and Sproull, 1959). As such, librarians have made, and continue to make, great efforts to preserve physical information. Analogous efforts have yet to coalesce to systematically preserve electronic information and collections, though several moderately dissociated initiatives exist, primarily for licensed content, through programs such as Portico, LOCKSS, CLOCKSS, among others (Bascones, 2012; Burnhill and Guy, 2010); and in the last few years these various archiving programs have started to see exponential growth in the amount of content being archived. Portico (http://www.portico.org/digital-preservation/), a not-for-profit, community-supported dark archive has entered into many new partnerships with publishers and reported one billion files archived in early 2016 (Pool, 2016). LOCKSS (https://www.lockss.org/), or Lots of Copies Keep Stuff Safe, is an open-source preservation system from Stanford University Libraries that allows an institution to preserve its own owned electronic content. CLOCKSS (https://www.clockss.org/clockss/Home), or Controlled LOCKSS, uses LOCKSS in a closed system, distributed dark archive. Both Portico and CLOCKSS provide hosted access to archived content after a trigger event; for example, a ceased publication that is no longer available, a publisher that has halted operations, or a catastrophe on the publisher's end that renders content unavailable. Unlike CLOCKSS and LOCKSS, Portico makes available triggered content to paid participating institutions. In a 2015 article, Mering presents an excellent overview of Portico, LOCKSS, CLOCKSS, and a few other archiving initiatives.

Over the years, librarians and publishers have worked together to establish digital archival methods and systems to ensure some sort of long-term access to (some) content (for some people) in perpetuity. Preservation and long-term access may be available if you can muster the understanding and interest of those able to pay for it (Pool, 2015). What if neither the understanding nor the funds to ensure long-term access to electronic collections exist? Is content at risk?

Conceivably, born-digital content that lacks a print analog might also lack permanence. Commercial publishers of scholarly journals have an imperative to secure long-term access for content, which is primarily bonded and funded through licensed subscriptions. OA publishers may not necessarily espouse a similar imperative, and OA electronic journals have been identified as content potentially at risk of loss or, in a best-case scenario, transience (Kenney and Wesley, 2016; Luther *et al.*, 2010; Regan, 2016; Worthey, 2009). Several analyses of OA journals found between one and four percent of titles not accessible online (Kaufman-Wills, 2005; Morris, 2006). In "Archiving in the networked world: Open access journals," Seadle (2011) found only eight percent of the titles indexed in the Directory of Open

Access Journals (DOAJ) were archived in LOCKSS/CLOCKSS and only five percent had been archived in Portico.

Research question

Problems accessing electronic journals are not uncommon, and they are often caused by OpenURL link resolution failures. OA electronic journals are no different in this regard, but occasionally OA titles vanish without notice. The disappearing acts are usually due to domain changes or a server hiccup, and the journals are eventually rediscovered. In 2015, a routine problem report uncovered an OA electronic journal that seemed to have completely disappeared.

A brief case of Gestalt!

An investigation of the reported access problem showed an OA title, *Gestalt!*, ceased publication and dematerialized. At the time, Ulrichsweb indexed the journal as electronic-only with a status of active. WorldCat reported 72 libraries worldwide holding the title. Google searches for the journal and the publisher did not provide additional information. A search using the Internet Archive Wayback Machine revealed a 2011 publisher change. The Wayback Machine also documented the slow decay of the URL between 2011 and 2015. The domain was taken over by an adult web site and then appears to have been sold to a domain name speculator.

The last reported publisher of the journal was contacted, after being found on web pages archived by the Internet Archive Wayback Machine. A representative for the presumed former publisher had no information about Gestalt!, nor could they find a current contact for the journal (McGilton, 2016). Some full-text content of the journal remains archived and presently accessible through the Wayback Machine.

The disappearance of *Gestalt!* prompted an initial attempt to answer the more significant question: are OA electronic journals really at risk? This study is a broad, preliminary investigation to test the current accessibility of OA electronic journals. It is by no means exhaustive or prescriptive, instead operating under the assumption that the accessibility or inaccessibility of a title could be used as a starting point to identify possible problems and to collect data for further analysis.

Method

The DOAJ is one of the most comprehensive international indexes for OA electronic journals, and its metadata is freely available for download in a comma-separated values file (Directory of Open Access Journals, 2016b). This metadata was downloaded, imported into an Excel file, and used to test journal accessibility. At the time of download, the file included 9,073 journals from 123 countries. To check for accessibility, three tests were run.

The first two tests were automated, employing a slightly modified Excel Visual Basic for Applications macro (Ross, 2012) to make a HTTP HEAD request of each journal URL and subsequently report the HTTP status code received. It is common to use this type of HTTP request to check URL validity, and Fielding and Reschke identified the HEAD method as "often used for testing hypertext links for validity, accessibility, and recent modification" (2014, p. 25). The two tests were identical and occurred approximately one month apart. This was done to compare overall results, in an effort to quickly and informally gauge the accuracy of the numbers found in the first test. It is quite possible that inaccessibility found during the first test was temporary, or it could have been found to be resolved in subsequent tests. It can be difficult to predict the instability and uncertainty of the internet on any given day.

A third manual test further analyzed the results of the first two tests. The different software or web servers used by journal publishers and providers may not adhere to identical practice, which could mean that some web sites may simply refuse to respond to an automated HTTP request. Therefore, inaccessible journals from the automated tests were combined into a single set of results. After title de-duplication, the URL of each journal was manually checked in a web browser.

Results

The full set of results for the two automated tests is summarized in Table I. The results show very similar numbers. Test one found that 69.51% of the HTTP requests made were successful, compared to 69.34% in the second test. Another 25.86% or 25.37% of the URLs were accessible after a redirect. Overall, 95.37% or 94.71% of the URLs tested for validity were accessible in some manner. Consequently, 4.63% or 5.29% of the URLs returned errors that indicated they were inaccessible; 420 or 480 titles, respectively.

As shown in Table II, HTTP error 404 – not found was the most common error among inaccessible URLs in both tests, accounting for a little more than one quarter of the errors. HTTP 404 errors can occur for a variety of reasons and may be temporary or permanent (Fielding *et al.*, 1999). The two tests differed slightly in the second most common error, and were again identical in the third and fourth most common errors. Both of these error codes, HTTP 403 – forbidden and HTTP 405 – method not allowed, may indicate a limitation in the data collection method. It is possible that the URLs may have been accessible, but the queried servers did not permit this specific type of automated HTTP request. The manual test addresses this possibility.

The inaccessible titles from both tests, and the DOAJ metadata associated with each title, were collated into a single set of results. Out of the combined list of 900 titles, 319 duplicates were identified.

The remaining 581 unique titles were tested a third time by manually checking each URL in Google Chrome. After the third test, 192 titles or 2.11% were still inaccessible. Of the 389 titles successfully reached on the third attempt, a little over 20% or 82 titles were accessible after a redirect. The DOAJ metadata was then examined in an attempt to uncover any common themes or patterns among the inaccessible titles. No solid patterns emerged. A complete list of the data fields is included in the appendix.

Publisher and journal data

The 192 inaccessible titles amount to 119 publishers. The majority of the publishers, 112 in all, have a single inaccessible title. Table III shows the seven publishers with more than one inaccessible title. Five publishers have two inaccessible titles, one publisher has three inaccessible titles, and one publisher stands out with 67 of the inaccessible titles. Of the 119 publishers, 27 of them indicated status as a society or institution. A mere seven publishers, constituting eight of the inaccessible titles, indicated that they participated in a digital archiving program or had an archiving policy in place. The 192 inaccessible journals are of a wide range of subjects, specifically 111 subjects and sub-classifications. As shown in Table IV, if only the main subject is counted, the most common subject is medicine followed by science, technology, social sciences, and education. None of the journals qualified for the DOAJ Seal, which directly correlates to the lack of archiving plans. Very few, only 36, of the journals even provided information on review processes.

Discussion

Searching for commonalities among the metadata that provides information about the journal and the publisher did not reveal a discernably obvious pattern. The biggest pattern, perhaps, was the pattern of little to no information. The DOAJ has already begun work to remedy this problem. In 2015 publishers were asked to commence a formal reapplication process to continue being indexed in the DOAJ. A 2016 blog post entitled "DOAJ to remove approximately 3300 journals," explained that the process had been initiated to "[ensure] that all journals in DOAJ…met the higher criteria for indexing that the DOAJ launched in March 2014" (Directory of Open Access Journals, 2016a). Among other things, the new criteria includes a strong recommendation for publisher archiving. Publishers may still apply and index their journals with the DOAJ without enacting a formal archiving plan; but to qualify for the DOAJ Seal, publishers must "have an archival arrangement in place with an external party for the long-term preservation and archiving of the journal's published content" (Olijhoek *et al.*, 2015, p. 5). The preservation organization Portico has reported receiving an uptick in business from OA journal publishers (Pool, 2015). It could turn out that several of these mildly disjointed concerns equate to good fortune for

the future sustainability and persistence of OA electronic journals; meaning that if a journal does unexpectedly disappear, it could very well continue to exist in a stable online archive.

A different issue presented itself with unexpected consistency in all three of the URL tests. Each time, regardless of the number of URLs being tested, roughly 20-25% of the URLs redirected to another URL. This may portend a future problem with OA electronic journal persistence and preservation; however, it also shows the complicated landscape of managing e-journals in the present, open access or not. In addition to the adoption of digital archiving programs, perhaps librarians and other supporters of the open access movement should encourage more OA publishers to endorse the Transfer Code of Practice (NISO, 2015) or adopt similar practice. Transfer focuses on tracking journals transferring between publishers or publishing platforms, and it provides title transfer alerts and maintains a database of transferred titles. This type of practice could be used to track OA electronic journals migrating between URLs.

Conclusion and future research

Open access publishing and electronic journal preservation are complex topics with lingering uncertainty. While this study found only 2.11% of 9,073 journals to be inaccessible, which may be a hearteningly small number, it prompts ideas for future research. If inaccessible titles have changed URLs or publishers, there is a chance they could be tracked down on the internet, in a union catalog or periodicals database, or by contacting the last known publisher directly. Additionally, at least one of the publishers of the inaccessible titles is included in Beall's List (https://scholarlyoa.com/publishers), which is a list of possible predatory open access publishers. This paper reserved a qualitative investigation into predatory publishers as it is an entirely different problem afflicting inaccessible or disappearing open access electronic journals.

References

Arnold, R.B. (1997), "The preservation of paper: How long will the page you are reading last?", *Serials Review*, Vol. 23 No. 4, pp. 15-26.

Barrow, W.J. and Sproull, R.C. (1959), "Permanence in book papers", *Science*, Vol. 129 No. 3356, pp. 1075-1084.

Bascones, M. (2012), "JISC collections: Post-cancellation entitlement registry scoping project", *Collaborative Librarianship*, Vol. 4 No. 3, pp. 85-95.

Beall, J. (2016), "Beall's list", available at https://scholarlyoa.com/publishers (accessed 3 October 2016).

Beh, E. and Smith, J. (2012), "Preserving the scholarly collection: An examination of the perpetual access clauses in the Texas A&M University Libraries' major e-journal licenses", *Serials Review*, Vol. 38 No. 4, pp. 235-242.

Bolter, J.D. (1987), "Text and technology: Reading and writing in the electronic age", *Library Resources and Technical Services*, Vol. 31 No. 1, pp. 12-23.

Burnhill, P. and Guy, F. (2010), "Piloting an E-journals Preservation Registry Service (PEPRS)", *The Serials Librarian*, Vol. 58 No. 1-4, pp. 117-126.

Directory of Open Access Journals. (2016a), "DOAJ to remove approximately 3300 journals", available at: https://doajournals.wordpress.com/2016/05/09/doaj-to-remove-approximately-3300-journals/ (accessed 8 August 2016).

Directory of Open Access Journals. (2016b), "How can I get journal metadata from DOAJ?", available at https://doaj.org/faq#metadata (accessed 24 June 2016).

Duranceau, E.F. (1998), "Archiving and perpetual access for web-based journals: A look at the issues and how five e-journal providers are addressing them", *Serials Review*, Vol. 24 No. 2, pp. 110-115.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masiner, L., Leach, P. and Berners-Lee, T. (1999), "Hypertext Transfer Protocol – HTTP/1.1", RFC 2616.

Fielding, R. and Reschke, J. (2014), "Hypertext Transfer Protocol (HTTP/1.1): Semantics and content", RFC 7231.

Kaufman-Wills (2005), "The facts about open access: A study of the financial and non-financial effects of alternative business models for scholarly journals", Association of Learned and Professional Society Publishers.

Kenney, A. and Wesley, K. (2016), "Building a social compact for preserving e-journals", *The Serials Librarian*, Vol. 70 no. 1-4, pp. 72-84.

Keyhani, A. (1998), "Creating an electronic archive", *The Serials Librarian*, Vol. 34 No. 1-2, pp. 213-224.

Leibowitz, F.R. (2002), "Serials staffing for the twenty-first century and beyond", *Serials Review*, Vol. 28 No. 3, pp. 195-200.

Luther, J., Difiore, K., Gibbs, N., Lamoureux, S.D., Reich, V., Staines, H.R. and Steinle, K. (2010), "Ensuring perpetual access to online subscriptions", *Serials Librarian*, Vol. 58 No. 1-4, pp. 73-78.

Mering, M. (2015), "Preserving electronic scholarship for the future: An overview of LOCKSS, CLOCKSS, Portico, CHORUS, and the Keepers Registry", *Serials Review*, Vol. 41 No. 4, pp. 260-265.

McGilton, M. (2016), private email, 6 August.

McMillan, G., Craighead, L.M., Dougherty, W.C., Kownacki, B., Metz, P., and Litchfield, C. (1991), "Electronic journals: Considerations for the present and the future", *Serials Review*, Vol. 17 No. 4, pp. 77-86.

Morris, S. (2000), "Archiving electronic publications: What are the problems and who should solve them?", *Serials Review*, Vol. 26 No. 3, pp. 64-68.

Morris, S. (2006), "When is a journal not a journal? A closer look at the DOAJ", *Learned Publishing*, Vol. 19 No. 1, pp. 73-76.

National Information Standards Organization. (2015), "Transfer code of practice", NISO RP-24-2015.

Olijhoek, T., Mitchell, D., and Bjørnshauge, L. (2015), "Criteria for open access and publishing", *ScienceOpen Research*, available at: https://www.scienceopen.com/document?id=85a98041-8734-4a43-b6eb-ff5903f3ae96.

Pool, R. (2016), "The rise and rise of digital preservation", *Research Information*, February/March, pp. 8-10.

Pool, R. (2015), "Preserving progress for future generations", *Research Information*, February/March, pp. 21-23.

Regan, S. (2016), "Strategies for expanding e-journal preservation", *The Serials Librarian*, Vol. 70 No. 1-4, pp. 89-99.

Ross, L. (2012), "ValidateURLs macro", available at: http://www.mrexcel.com/forum/excel-questions/632198-check-if-url-valid.html (accessed 24 June 2016).

Rothenberg, J. (1995), "Ensuring the longevity of digital documents", *Scientific American*, Vol. 272 No. 2, pp. 42-47.

Seadle, M. (2011), "Archiving in the networked world: Open access journals", *Library Hi Tech*, Vol. 29 No. 2, pp. 394-404.

Szydlowski, N. (2010), "Archiving the web: It's going to have to be a group effort", *The Serials Librarian*, Vol. 59 No. 1, pp. 35-39.

Worthey, G. (2009), "'Long live the corpse!' At risk open-access humanities journals in LOCKSS", *Against The Grain*, Vol. 21 No. 1, pp. 22-26.

Journal title	First calendar year journal provided online Open Access content
Journal URL	Full text formats
Alternative title	Keywords
Journal ISSN (print version)	Full text language
Journal EISSN (online version)	URL for the Editorial Board page
Publisher	Review process
Society or institution	Review process information URL
Platform, host or aggregator	URL for journal's aims & scope
Country of publisher	URL for journal's instructions for authors
Journal article processing charges (APCs)	Journal plagiarism screening policy
APC information URL	Plagiarism information URL
APC amount	Average number of weeks between submission and publication
Currency	URL for journal's Open Access statement
Journal article submission fee	Machine-readable CC licensing information embedded or displayed in articles
Submission fee URL	URL to an example page with embedded licensing information
Submission fee amount	Journal license
Submission fee currency	License attributes
Number of articles publish in the last calendar year	URL for license terms
Number of articles information URL	Does this journal allow unrestricted reuse in compliance with BOAI?
Journal waiver policy (for developing country authors etc)	Deposit policy directory
Waiver policy information URL	Author holds copyright without restrictions
Digital archiving policy or program(s)	Copyright information URL
Archiving: national library	Author holds publishing rights without restrictions
Archiving: other	Publishing rights information URL
Archiving infomation URL	DOAJ Seal
Journal full-text crawl permission	Tick: Accepted after March 2014
Permanent article identifiers	Added on Date
Journal provides download statistics	Subjects
Download statistics information URL	

Table AI.

DOAJ metadata fields

| | Test one | | Test two | |
Status code	Number	%	Number	%
HTTP 2xx - success	6,307	69.51	6,291	69.34
HTTP 3xx - redirection	2,346	25.86	2,302	25.37
HTTP 4xx - client error	213	2.35	214	2.36
HTTP 5xx - server error	47	0.52	63	0.69
Unspecified/other	160	1.76	203	2.24
Total	9,073	100.00	9,073	100.00

Table I.

Summary count of status codes returned

| | Test one | | Test two | |
Error	Number	%	Number	%
HTTP 403	49	11.67	53	11.04
HTTP 404	112	26.67	107	22.29
HTTP 405	47	11.19	47	9.79
HTTP 406	4	0.95	6	1.25
HTTP 416	1	0.24	1	0.21
HTTP 500	31	7.38	41	8.54
HTTP 502	2	0.48	3	0.63
HTTP 503	13	3.10	17	3.54
HTTP 504	1	0.24	0	0.00
HTTP 522	0	0.00	1	0.21
HTTP 523	0	0.00	1	0.21
A connection with the server could not be established	6	1.43	8	1.67
The connection with the server was terminated abnormally	8	1.90	5	1.04
The host name in the certificate is invalid or does not match	0	0.00	1	0.21
The operation timed out	43	10.24	97	20.21
The server name or address could not be resolved	50	11.90	41	8.54
The server returned an invalid or unrecognized response	27	6.43	25	5.21
The URL is invalid	26	6.19	26	5.42
Total	420	100.00	480	100.00

Table II.

Count of error codes returned

Publisher	Number	
Bentham open	67	
Universidade de BrasÃlia	3	**Table III.**
Universidad de Cienfuegos	2	
Universidad Nacional de Colombia	2	
Polish Academy of Sciences	2	Count of titles by
De Gruyter Open	2	publishers with
SOLON	2	more than one
Total	80	inaccessible title

Subject	Number	
Medicine	49	
Science	30	
Technology	19	
Social Sciences	18	
Education	17	
General Works	13	
Agriculture	10	
Language and Literature	10	
History (General) and history of Europe	7	
Geography	5	
Philosophy	5	
Law	3	**Table IV.**
Bibliography	3	
Political science	2	Count of
Military Science	1	inaccessible
		titles by primary
Total	192	subject