The growth and decay of URLs citation: A case of an online Library & Information Science journal

Sumeer Gul, Iram Mahajan and **Asifa Ali**
Department of Library and Information Science,
University of Kashmir, INDIA
e-mail: sumeersuheel@gmail.com (corresponding author)

ABSTRACT

The present study makes an attempt to explore and analyze the growth and decay rate of URL citations cited in one of the eminent information web magazine ARIADNE. All the web citations cited in the featured article section of ARIADNE web magazine spanning a period of three years (2010-2012) were identified and downloaded. All the URLs provided in the web citations were checked by W3C link checker (http://validator.w3.org/checklink) to test whether they were accessible or not. Web resources are used more than the printed ones. Though the number of references in web based format is more in use but a significant number of web references lose their identity with the passage of time. The results indicate that early published papers have collectively a larger number of missing web citation compared to the recent ones. It was found that majority of errors were due to the missing content (http 404-file not found) representing 52.68% of all http error codes followed by "http 500" (24.73%) and "http 403" (19.35%). The ".com/.co" domain was also found to be the most stable and persistent domain with 95 per cent accessibility. The greatest number of web resources cited in the articles, were found to be of "html" and "htmls" formats and "ppt" files were found to be most stable with 100% accessibility. Finally, the study confirms that continued availability of web resources is not guaranteed because of the dying phenomenon of web based references, but there can be certain solutions (WebCite, LOCKSS etc) that can prevent the decay or disappearance of web citations.

Keywords: **Web citations;** Missing links; Dead URL's; URL persistency; Broken links; URL Accessibility.

INTRODUCTION

The Web is one of the most imperative and multifaceted innovations in human history. It is the largest and most complete tool for information exchange ever made accessible to the global population. It has turned as an information rivulet crossing all borders. Millions of users access the Web every day for diverse beneficial purposes ranging from hobbies and communications to research and learning. With the development of the Web, several scientific journals are available in electronic form for researchers who had made the Web one of the main communicational tools. The accessibility options of researchers are changing drastically from print based medium to the online one. Accordingly, citations from web resources (URLs) as novel references that form a part of electronic journals have increased. The main reason for the increase in web citations in scholarly papers is that the Web has become the foremost preference for finding information, such as current research, making scientific discoveries and keeping up with colleagues at other institutions. The extensive adoption of the Web as a mechanism for sharing information has brought

with it the corresponding ubiquity of URL references and citations. The World Wide Web is growing at a gargantuan speed and web citations are becoming very frequent in scholarly publications. Citation has several important purposes: to avoid plagiarism, to attribute prior or unoriginal work and ideas to the correct sources, to allow the reader to determine independently whether the referenced material supports the author's argument in the claimed way, and to help the reader gauge the strength and validity of the material the author has used. The citing behavior of researchers has been influenced by the Web and this has subsequently influenced the growth of web citations (Moghaddam, Saberi and Esmaeel 2010)

Casserly and Bird (2008) stated that students and researchers look to literature citations as links between what is new and what is already known but the foremost problem of web citation is their persistence, since citations disappear over time, causing accessibility problems for readers. Rumsey (2002) suggested that with the increase in scholarly citations to web-located resources, there is a need to be also be cautious due to the way that URL references tend to disappear. The Web is growing at an enormous speed, and has become an indispensable source of information for research. New pages are constantly added, but there are additional processes as well: pages are moved or removed and/or their content changes (Barr-Ilan and Peritz 2009).This concern with disappearing web-resources is manifested through the growing incidence of *"missing"* or *"broken"* or *"dead"* web links. The valuable resources tend to disappear over time, thus depriving the users of the information.

Very often even online archives or search engines are unable to bring up missing online information. Spinellis (2003) stated that there are 24 different errors that can occur within an *"HTTP exchange"*. In addition, some errors can occur before the communication between the client and server take place. In practice, whenever an URL is inaccessible an error message (HTTP code) appears. Lawrence et al. (2001) through manual analysis of invalid links identified a number of reasons why URLs become invalid: (a) personal homepages are likely to disappear with the change in location of researcher, hence changing their home page URLs, (b) restructuring of sites without maintaining the old links, (c) changing the protocol portion of the URL from FTP servers to HTTP.

The use of web citations in the scholarly publications is becoming more common with the passage of time as the Web is growing at massive speed. But due to the ever changing, instable and secular identity of the Internet, web citations (URLs) are exposed to the risk of rot. The present study makes an attempt to explore and analyze the growth and decay rate of URL citations cited in one of the eminent Information Science web magazine "Ariadne".

ABOUT ARIADNE

Ariadne is a web magazine for information professionals in higher education based in the UK and worldwide. Ariadne is provided by the University of Bath Library to inform policy-makers and practitioners in higher education libraries and associated sectors of developments in the online environment. Since its launch in January 1996 by UKOLN (formerly known as the United Kingdom Office for Library and Information Networking), it has attempted to keep the library practitioners abreast of current digital library initiatives, as well as technological developments. It is focused on in-depth reporting, to the information community at large, on the progress and developments within the UK Electronic Libraries Programme (eLib), covering matters such as information service

developments and information networking issues worldwide. Ariadne generally seeks to provide readers with a variety of articles in each issue, some technical, some of a more strategic nature, in the anticipation that most readers will find something of relevance to their work or interests in every issue (About Ariadne 2013).

REVIEW OF RELATED LITERATURE

Researchers today prefer to use web resources over print because the ever-increasing number of scholarly materials available on the Web makes their job of referencing easier. The trend of using web citations has become common in scholarly publications as it has become convenient to hyperlink between documents, images, audio, and video resources (Goh and Ng 2007), but studies have shown that up to 23% of Uniform Resource Locator (URL) references became inactive after 1 year and up to 52% after 5 years (Lawrence et al. 2001; Dellavalle et al. 2003; Spinellis 2003; Wren 2004). When any information is taken from the Internet, the URL of that source is provided so that readers can have access to the original content and get to know where that particular work is hosted on the Web (Nagaraja et al. 2011). Web citations face a number of problems regarding the persistence and availability of URLS. Several researchers have studied the instability and decay of web references and the related issues (Koehler 1999; Koehler 2002; Koehler 2004; Markwell and Brooks 2002; Markwell and Brooks 2003; McCown et al. 2005; Moghaddam, Saberi and Esmaeel 2010; Rumsey 2002; Sellito 2004; Spinellis 2003; Wagner et al. 2009).

The availability and permanence of URLs in electronic journals and scholarly communications is also reported through many studies (Harter and Kim 1996; Germain 2000; Davis and Cohen 2001; Carnevale and Aronsky 2007; Tajeddini et al. 2011). The URL inactiveness has also been studied by Koehler (1999). According to Ducut, Liu and Fontelo (2008), URLs decay for several reasons: server shutdowns; reconfiguration or content change of URLs; error in citing a URL; removal of a website and content change or redirection. Nielsen (2008) is of the opinion that "Linkrot contributes to dissolving the very fabric of the web: there is a looming danger that the web will stop being an interconnected universal hypertext and turn into a set of isolated info – islands". Dimitrova and Bugeja (2007) found that 10 to 40% of web-based citations are likely to disappear from the Web. In a study on *"linkrot"* Casserly and Bird (2003) found an availability rate of web references declined from 61.80 to 30.27 %. A study conducted by Ducut, Liu and Fontelo (2008) reviewing website accessibility three years after it was initially accessed, showed over 2/3 of URLs could not be found anymore or had moved with no forwarding URLs. They also reported that only 78% of URLs were generally available at the time of access. Goh and Ng (2007) found that not many articles cited other online articles. Even, about half of those references were inaccessible later in the same year. Their results also showed that links are subject to a high decay rate. Recent study placed URL accessibility at about 70% with an annual decay rate of 5.4%. Veronin (as cited in Kumar and Raj 2012) examined a very high rate of *"URL decay"*. Kumar and Raj (2012) also found that websites, in general, have a comparatively short lifespan. URL sturning victims of *"link rot,"* becoming non-viable, moving without automatic forwarding, or having their content changed have also been studied by Markwell and Brooks (2003).

Wren (2004) found that *"URL decay"* rate was higher within full text documents ranging from 30 % to 60 %. Nagaraja et al. (2011) observed that *"URL decay"* rate was increasingly more definite as time elapsed. A difference was also observed by article type as a lower rate of attrition occurred in research articles (14.5 per cent) compared to non-research

articles (18.0 per cent). They also found that Error *"404-page not found"* was the most common error occurred across all three years, accounting for 84.9% of outdated URLs. Sellitto (2005) found that some 46% of all citations to web sources could not be accessed with the *HTTP 404 ("Page not found")* message being the supreme cause of missing citations. Collectively, 22% of all citations were missing, which represents a significant reduction in the theoretical knowledge base underpinning loads of scholarly articles. Disappearance of web-located citations diminishes the opportunities of future citation. Spinellis (2003) revealed that successful retrieval rates differ depending on the URL source. The most common reason for inaccessibility of URLs is that the resource referenced has been removed from the server ("Error 404") followed by invalid host name ("Error-901") and a small representation is also because of network problems ("Error-504"). Saberi and Abedi (2012) are of the view that error messages *"403"* and *"401"* are possibly due to the firewall software, filtering and Internet limitations and other messages *"404", "500" [. . .]* are possibly due to the deletion of web site or transfer of data to other web sites. According to Gomes and Silva (2005), on an average 0.8% of URL death is identified through *"HTTP"* errors. The most common *"HTTP"* errors are *"File Not Found" (404)*, *"Internal Server Error" (500)* and *"permanent or temporary redirections" (301, 302)*. The main causes of URL decay are the frequent substitution of URLs and site death.

Many studies have estimated the "half-life period" (time required for half of all online citations in journal to disintegrate) of web citations in various disciplines and researchers found that this amount of time may differ (Saberi and Abedi 2012; Moghaddam, Saberi and Esmaeel 2010). Goh and Ng (2007) investigated the link decay phenomenon in three leading Information Science journals. Articles spanning a period of seven years (1997-2003) were downloaded and their links were extracted. From these, a measure of link decay, the half life, was computed to be approximately 5 years. Kumar and Kumar (2012) found the half-life of URL citations cited in LIS scholarly journals was approximately 11.5 years. Moghaddam and Saberi (2011) observed that as the age of URL citations increases, the number of inaccessible URLs increases.

OBJECTIVES AND METHOD

The study was conducted for the purpose of studying the status and rate of web citations of articles published in Ariadne with regards to:
   a) Distribution of web and print citations by year
   b) Distribution of active/dead URL citations
   c) Error codes associated with missing URL'S
   d) Domains associated with missing links
   e) Distribution of missing URLS by file formats
   f) URL scheme distribution among missing web citations

All the web citations cited in the articles of *"featured article section"* of Ariadne spanning over a period of three years (2010-2012) were identified and downloaded directly from the web site and checked in terms of decay and availability of individual cited URLs. Editorial, tooled up, book review and event report sections were omitted. URLs provided in the reference section of featured articles were identified and examined. Total of 1184 web citations were recorded from 73 articles. All the URLs provided in the web citations were checked by *"W3C link checker"* (http://validator.w3.org/checklink)[1] to test whether they

---

[1] Link checker tests the submitted URLs for broken or non-valid hypertext links and reports the errors encountered. World Wide Web consortium (W3C) link checker is freely available on web.

were accessible or not. URLs that were reported inaccessible or broken, were further checked after 10 days, as such URLs were found to be active. Similar procedure was repeated after every 15 days in the three month (March 2013 to May 2013) to determine whether the cited URLs still exist. Final result of these tests were recorded and tabulated in MS-Excel spreadsheet. URL citations with HTTP errors, domains, URL scheme and file type associated with missing URL citations were listed separately for further statistical analysis.

RESULTS

Distribution of Web and Print Citations by Year

A total of 1392 citations were identified from 73 articles out of which 1184 were web citations and 208 were print. Table 1 presents the distribution of web and print citations by year. The table also shows the average number of URL citations per article, which ranged from a maximum of 17.52 in 2010 to a minimum of 15.17 in 2012, as the number of articles published in the year 2010 was more than rest of the years. The average number of web citations was 16.21 across all articles and years. The web citation as a percentage of all citation also significantly increased from the year 2010 to 2012 (83.82 in 2010 to 87.68 in 2012).

Table1: Distribution of Web and Print Citations in Ariadne by year

Year	No. of Articles	Total Citations	Web Citations	Average web Citations per Article	% age	Print Citations	% age
2010	34	711	596	17.52	83.82	115	16.17
2011	16	283	239	14.93	84.45	44	15.54
2012	23	398	349	15.17	87.68	49	12.31
Total	73	1392	1184	16.21	85.05	208	14.94

(Percentage is calculated on the total number of citations per year)
Distribution of active/dead URL citations

Table 2 shows a total of 1184 web citations, out of which 1091 (92.14%) URLs were active and 93 (7.85%) URLs were missing. The maximum number of missing URLs are encountered during the year 2010 (58), followed by 2011 (20) and 2012 (15). Table 2 also shows the decline in number of missing web citations moving from the year 2010 (9.73%) to 2012 (4.29%).

Table 2: Summary of Active/Missing Web Citations

Year	Total web citations	Active citations	% age of active URLs	Missing citations	% age of missing URLs
2010	596	538	90.26	58	9.73
2011	239	219	91.63	20	8.36
2012	349	334	95.70	15	4.29
Total	1184	1091	92.14	93	7.85

Error Codes Associated with Missing URL'S

Five types of error codes associated with missing URL citations were encountered during the study (Table 3). "*HTTP 404-page not found*" is encountered in majority of missing URLs representing 52.68% of all *"HTTP errors"*. A total of 24.73% missing URL citations shows "HTTP 500-internet server error", followed by "*HTTP 403-forbidden*" (19.35%), "*HTTP 503-service unavailable*" (2.15%) and "*HTTP 400-bad request*" (1.07%).

Table 3: HTTP Error Codes Associated with Missing URLs for the Year 2010-2012

Year	HTTP 400	HTTP 403	HTTP 404	HTTP 500	HTTP 503	Total
2010	0	9	35	14	0	58
2011	1	6	5	6	2	20
2012	0	3	9	3	0	15
Total	1	18	49	23	2	93
% age	1.07	19.35	52.68	24.73	2.15	

Domains Associated with Missing Links

Five different types of domains were taken into account from the accessibility and decay viewpoint involving *".org",".edu/", ".ac",".com",".gov" and ".net"*. Nevertheless, those domains not belonging to any of these categories were categorized in *"others"* (Table 4). Domains of the cited URLs mostly include the *".edu/.ac"* type. Accordingly, from 1184 web citations, the highest number of domains (376) is of*".edu/.ac"* type, followed by *".com/.co"* type (341). There were very less number of URLs with *".gov"* domain (55). Moreover, the percentage of accessibility and decay of domains proves that very large percentage of URLs of *".gov"* domain (16%) have acquired decayed followed by *".net"* (11%), *".edu/.ac"* (9%), *".ac"* (7%) and *".com/.co"* (5%).

Table 4: Distribution of URL Citations by Domain Type

Domains	Total URLS	Accessible	% age	Inaccessible	% age
.com/.co	341	324	95	17	5
.edu/.ac	376	342	91	34	9
.net	61	54	89	7	11
.org	290	271	93	19	7
.gov	55	46	84	9	16
others	61	54	89	7	11
Total	1184	1091	92	93	8

Distribution of Missing URLS by File Formats

Table 5 indicates that the most number of cited web resources are *"html"* based files. Out of 1184 web citations, 824 cases are in *"html"* file format, followed by 108 *"pdf"* files, 60 *"php"* files, 16 *"ppt"* files and 9 *"htmls'* files. A total of 108 file types were included in the *"other"* category. It is also evident that none of the *"htmls"* and*"ppt"* files got decayed or missing. However, a large number of *pdf* files (13% of total pdf files) decayed.

Table 5: Distribution of Active and Dead URLs by File Format

File format	Total URLS	Active URLS	% age	Dead URLS	%age
html	883	824	93	59	7
htmls	9	9	100	0	0
pdf	108	94	87	14	13
php	60	53	88	7	12
ppt	16	16	100	0	0
others	108	95	88	13	12
Total	**1184**	**1091**	**92**	**93**	**8**

URL Scheme Distribution among Missing Web Citations

Table 6 shows the distribution of schemes in the set of 1184 URL citations. A total of 1143 URLs relied on *"http"* scheme, out of which 1054 (92.21%) were accessible and 89 (7.78%) were inaccessible. Very small number of URLs (41) relied on *"https"* scheme, out of which 37 (90.24%) were accessible and 4 (9.75%) were inaccessible.

Table 6: URL Scheme Distribution

Scheme	http	https
Accessible	1054	37
Inaccessible	89	4
Inaccessible (%age)	95.7	4.3

Table 7 shows the distribution of URL scheme among the inaccessible URLs. An overwhelming majority of the URLs using the *"http"*scheme (95.70%) of total inaccessible URLs were missing, whereas a small number (4.30%) of missing links use*"https"* scheme. However, none of the recorded missing web citations used *"ftp", "gopher"* and *"file"* scheme.

Table7: Percentage Distribution of Missing URLs by Year

Year	http				https			
	Active	% age	Dead	% age	Active	% age	Dead	% age
2010	525	49.81	56	62.92	13	35.13	2	50
2011	212	20.12	19	21.35	7	18.91	1	25
2012	317	30.07	14	15.73	17	45.94	1	25
Total	**1054**	**92.21**	**89**	**7.78**	**37**	**90.24**	**4**	**9.75**

DISCUSSION AND CONCLUSION

The Web is becoming the first choice among scholars for identifying valuable information like current research, new inventions or discoveries etc. The Web has influenced the citing behavior of scientists all over the globe. On the other hand, with the growth of information

resources in the era of web the accessibility and persistence of web resources is a serious issue that is growing in significance. Reconstruction, terminating, merging, redirecting and expanding web sites can mean an inconsistency in web URLs. This phenomenon has been studied as web site persistence and decay (Dimitrova and Bugeja 2007; Moosavi, Moghaddam and Tajeddini 2012), web site constancy and permanence (Tyler and McNeil 2003), and web site accessibility and persistence of URLs. Kumar and Kumar (2012) wrote that web resources have undoubtedly changed information access but this also created challenges because of the decaying nature of URL citations. Spinellis (2003) also revealed that mobility and inaccessibility of web-based sources are still serious reasons that may affect scholar's tendency to use them as legitimate media of formal scholarly communication. The controversial issue of missing citations is growing at an increasing rate over time. URL decay or missing link phenomenon can be attributed to a number of reasons which has been discussed under the following major findings:

1) Citation distribution: A total of 73 articles were published from the year 2010-2012. A total of 1392 citations were identified from these articles, out of which 1184 (85.05%) are web citations and rest 208 (14.94%) were print citations which is indicative of more web based citation usage. Web citations cover the major portion of reference section in articles. Kumar and Raj (2012) and Lawrence et al. (2001) in their work also showed a substantial increase in the number of URL citations since the inception of the Web. Kumar and Raj (2012) even stated that web citations in the scholarly articles increased approximately six fold. This clearly shows a greater use of web resources by scholars than printed sources**.**

2) Rising trend of using web citation: The highest number of articles (34) was published during the year 2010, followed by 23 in 2012 and 16 in 2011. The average number of Web citations was 16.21 across all articles and years. Web citations exhibit an exponential increase from 2010 to 2012.The rate of use of web citations has increased from 83.82% in 2010 to 87.68% in 2012. Based on this result, it can be said that researchers are increasingly making use of Web-based resources. Kumar and Raj (2012) in their work revealed that the average number of web citation per article has increased from 0.25 (in the year 2001) to 3.02 (in the year 2008). Kumar and Kumar (2012) also in their study showed the significant increase in web citations per article. Percentage of URL citations per article increased from 5.25 percent in 1998 to 27.79 percent in 2009.

3) Distribution of active/missing URLs and age of citations: A total of 1184 web citations were examined and extracted URLs were checked. 1091 (92.14%) URLs were found to be active and 93 (7.85%) URLs were broken / missing. The percentage of missing links decreased from 9.73% in 2010 to 4.29% in 2012. The results indicate that early published papers have collectively a larger number of missing web citation compared to the recent ones. This clearly indicates that there is an association between the accessibility of URL citations and the age of URL citations. Ducut, Liu and Fontelo (2008), Goh and Ng (2007), Koehler (1999) and Spinellis (2003) in their work also observed that the URLs from recent years were more accessible than earlier ones. Kumar and Kumar (2012) stated as the age of the URL citations increase, the percentage of inactive URL citations tends to increase. Supporting Kumar and Kumar (2012), Aronsky et al. (2007 also observed inaccessibility of URL's within two days after an article's release to the public.

4) Error codes associated with missing links: The study also investigated types of link accessibility errors encountered. It was found that majority of errors were due to the missing content *(http 404-file not found)* representing 52.68% of all *"http"* error codes followed by *"http 500"* (24.73%) and *"http 403"* (19.35). Goh and Ng (2007), Kumar and Raj (2012), Kumar and Kumar (2013), Moghaddam and Saberi

(2011), and Moosavi, Moghaddam and Tajeddini (2012) and were also of the view that *"Error 404-file not found"* is the most overwhelming error message encountered. *"Error 404 (page not found)"* and *"Error 505 (internal server error)"* have also been attributed to non-availability of URL's (McCown et al. 2005).The reason for this is that there may be a link on the site that was wrong or the page might have been recently removed from the site. As there is no web page to display, the web server sends a page that simply says *"404 Page not found"* (404 Error Pages.com, n.d).

5) Domains associated with missing links: The domains of the cited URLs mostly include the *".edu/.ac"* type. From 1184 web citations, the highest numbers of domains (376) are of*".edu/.ac"* type. This reveals that the data sources of most of the web citations are web sites of various educational and academic organizations. Moreover, the percentage of accessibility and decay of domains, proved that*".com/.co"* with 95 percent accessibility is the most stable and persistent domain, while the*".gov"* domain with merely 84 per cent accessibility is the least stable. Surprisingly, the results for the*".gov"* domain were dissimilar from findings of other studies (Dimitrova and Bugeja 2006; Maharana, Nayak and Sahu 2006). However, there are other studies that have shown an immense rate of inaccessible URLs in the ".gov" domain. Casserly and Bird 2003; Markwell and Brooks (2003, as cited in Rhodes 2010), and Wagner et al. (2009) have shown that resources published at .gov URLs have a greater rate of link rot than those from other top-level domains. It was also recognized that the dominant domains in the study i.e. *".com/.co"* (5%) and *".org"* (7%) has less ratio of missing links. This result has been supported by Kumar and Kumar (2012) who were also of the view that *".com/.co"* and *".org"*domains are more stable. Wren (2008) has also observed that URLs published by organizations tend to be more stable.

6) File formats of missing URLs: The most numbers of cited web resources are *"html"* files (883). It has been found that *"htmls"* and *"ppt"* files are the most stable files than other type of files with 100 percent accessibility followed by *"html"* files with merely 7 percent decay. The highest percentage missing URLs were found in *"pdf"* (13%) followed by *"php"* (12%). The results for the *"pdf"* format were different from findings from most of other studies (Moosavi, Moghaddam and Tajeddini 2012; Saberi and Abedi 2012). It is difficult to determine if a relationship exists between file format type and link rot. Rhodes (2010) in an analysis of sample in 2008, 2009, and 2010 showed an increase in link rot among *"pdfs"* to 27% in 2010 from 8.2% in 2008. The same sample showed an increase in link rot among *"x/html"* web pages to nearly 35% in 2010 from 8.7% in 2008. This shows that there is no relationship between the type of the file format and link decay.

7) Distribution pattern of URLs of missing web citations: The distribution of schemes in the unique set of 1184 URLs showed an overwhelming majority of the URLs used the 'http' scheme (96.53%), whereas a small minority (3.46%) relied on https scheme. McCown et al. (2005) in their work also found the majority of URLs used the "http" scheme.

HTTP files were found to be more stable than https files with 92.2% accessibility and 7.78% decay, whereas as 9.75% of total htmls files were decayed making 90.24% of total htmls files accessible. HTTP files being more in number constitute the 95.7% of total missing links. Although authors may be aware of the dangers of using web citations, such as their decay and inaccessibility in the future, yet they cannot easily overlook the web resources in their articles, because the information they might be looking for could not be found in print (Falagas, Karveli and Tritsaroli 2007).

Neither publishers nor authors are able to guarantee the long-term precision or availability of online information because of the unpredictable nature of URLs. Effective solutions will likely require a collaborative effort on the part of researchers, authors and journal editors. Subsequently, several approaches of preservation of web content published in scholarly journals have been proposed, whether as policies and procedures (Johnson et al. 2004) or computationally, such as software tools (Eysenbach 2005; Kahle 1997; Reich and Rosenthal 2004; Schafer, Weibel and Jul 2001) and unique tagging/tracking measures like Digital Object Identifiers (DOI), Lots of Copies Keeps Stuff Safe (LOCKSS), and Persistent Uniform Resource Locator (PURL) (Caplan 1998).The LOCKSS program allows journals to partner with university libraries that keep a backup for any URL references that are published. DOIs allocate a permanent identifier to electronic resources and PURLs provide users with a transparent URL access mechanism that takes the user directly to the active location of required resource. The aforementioned cases can help increase accessibility and decrease the deterioration of web citations, but some researchers such as Wagner et al. (2009) believe that utilizing archiving tools such as WebCite is the best and only way to prevent web citations from deterioration and ensure that the cited web contents are accessible to readers in the future. However, authors should be cautious so as not to rely exclusively on these archiving tools.

The study confirms that continued availability of web resources is not guaranteed because of the dying phenomenon of web-based references. Though web-based resources are increasingly used in the research world but reliability on the persistence of web based resources can hardly be merited. Preserving the URL content is in need of immediate policies and plans to ensure persistence and longevity of web-based resources to save the research world and the researchers from the rotten content. New strategies to manage Internet resources and improved citation practices are necessary to minimize the future loss of information (Lawrence et al. 2001). Publishers need to adopt new policies to address information management challenges not previously encountered with printed literature. Finally, the study confirms that continued availability of web resources is not guaranteed because of the dying phenomenon of web based references, but there can be certain solutions (such as WebCite and LOCKSS) that can prevent the decay or disappearance of web citations.

ACKNOWLEDGEMENT

REFERENCES

404 - Error Pages.com. Available at: http://www.404errorpages.com

About Ariadne. (2013). Available at: http://www.ariadne.ac.uk/about

Aronsky, D., Madani, S., Carnevale, R. J., Duda, S., and Feyder, M. T. 2007. The prevalence and inaccessibility of Internet references in the biomedical literature at the time of publication. *Journal of the American Medical Informatics Association :Jamia*. Vol. 14, no.2: 232-234.

Bar-Ilan, J., and Peritz, B.C. 2009. The lifespan of "informetrics" on the Web: An eight year study (1998–2006). *Scientometrics*. Vol. 79, no. 1: 7-25.

Caplan, P. 1998. Do I or don't we? *PACS Review*. Available at: http://www.pr/v9/n1/capl9n1.html.

Carnevale, R. J., and Aronsky, D. 2007.The life and death of URLs in five biomedical informatics journals. *International Journal of Medical Informatics*, Vol. 76, no. 4: 269-273.

Casserly, M.F. and Bird, J.E. 2008. Web citation availability: A follow-up study. *Library Resources & Technical Services,* Vol. 52, no.1: 42-53.

Casserly, M. F. and Bird, J. E. 2003. Web citation availability: Analysis and implications for scholarship. *College and Research Libraries*, Vol. 64, no. 7: 301.

Davis, P.M., and Cohen, S.A. 2001.The effect of the Web on undergraduate citation behavior 1996-1999. *Journal of the American Society for Information Science and Technology*. Vol. 52, no.4: 309-14.

Dimitrova, D. V. and Bugeja, M. 2006. Consider the source: Predictors of online citation permanence in communication journals. *Portal: Libraries and the Academy*, Vol. 6, no.3: 280.

Dimitrova, D.V. and Bugeja, M. 2007. Raising the dead: Recovery the decayed online citations. *American Communication Journal*, Vol. 9, no. 2: 2.

Dimitrova, D.V., and Bugeja, M. 2007. The half-life of internet references cited in communication journals. *New Media & Society*, Vol. 9, no.9: 811-26.

Ducut, E., Liu, F. and Fontelo, P. 2008. An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation**.** *BMC Medical Informatics and Decision Making,* Vol. 8, no. 23: 2.

Eysenbach, G. 2005. Going, going, still there: Using the WebCite service to permanently archive cited web pages. *Journal of Medical Internet Resources*, Vol. 7, no.5: 60- 68.

Falagas, M.E., Karveli, E.A. and Tritsaroli, V.I. 2007. The risk of using the Internet as reference resource: a comparative study. *International Journal of Medical Informatics*, Vol. 77, no. 4: 280-286.

Germain, C.A. 2000. URLs: uniform resource locators or unreliable resource locators. *College and Research Libraries.* Vol. 61, no.4: 359-65.

Goh, D.H.L., and Ng, P.K. 2007.Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, Vol. 58, no.1: 17. DOI:10.1002/asi.20513

Gomes, D. and Silva, J. M. 2005. *On URL and Content Persistence*. Available at: www.xldb.fc.ul.pt/daniel/docs/papers/urlPersistenceRT.pdf

Harter, S.P., and Kim, H.J. 1996. Electronic journals and scholarly communication: a citation and reference study. *Information Research*, Vol. 2, no. 1. Available at: http://www.informationr.net/ir/2-1/paper9a.html

Johnson, K.R., Hester,E. J., Schilling, L.M. and Dellavalle, R.P. 2004. Addressing Internet reference loss. *Lancet*, Vol.363, no.9409: 660-661.

Kahle, B. 1997. Preserving the Internet. *Scientific American*, Vol. 276, no. 3: 72-74.

Koehler, W. 1999. An analysis of Web page and Web site constancy and permanence. *Journal of the American Society for Information Science*, Vol. 50, no. 2: 162–180.

Koehler, W. 2002. Web page change and persistence: a four-year longitudinal study. *Journal of the American Society of Information Science and Technology*. Vol. 53, no.2: 162-71.

Koehler, W. 2004. A longitudinal study of web pages continued: a report after six years. *Information Research*. Vol. 9, no. 2. Available at: http://InformationR.net/ir/9-2/paper174.html

Kumar, B.T.S. and Kumar, K.S.M. 2012. Decay and half life period of online: Citations cited in open access journals. *The International Information and Library Review*, Vol. 44, no. 4: 202-211.

Kumar, B.T.S., and Kumar, D.V. 2013. HTTP 404-page (not) found: Recovery of decayed URL citations. *Journal of Informetrics*, Vol.7, no.1: 145-157.

Kumar, B.T.S., and Raj, K.R.P. 2012. Availability and persistence of web citations in Indian LIS literature. *The electronic library*, Vol. 30, no.1: 22.

Lawrence, S., Pennock, D.M. , Flake, G.W. , Krovetz, R. , Coetzee, F.M. , Glover, E. ; Nielsen, F.A., Kruger, A. and Giles, C.L. 2001. Persistence of web references in scientific research, *IEEE Computers*. Vol. 34 no. 2: 26–31.

Maharana, B., Nayak, K. and Sahu, N.K. 2006. Scholarly use of web resources in LIS research: a citation analysis*. Library Review*. Vol. 55, no. 9: 599.

Markwell, J., and Brooks, D.W. 2002. Broken links: the ephemeral nature of educational WWW hyperlinks. *Journal of Science Education and Technology*. Vol. 11, no.2: 105-8.

Markwell, J. and Brooks, D. W. 2003. Link rot limits the usefulness of web-based educational materials in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*. Vol. 31, no. 1: 69–72.

McCown, F., Chan, S., Nelson, M.L and Bollen, J. 2005.The Availability and persistence of web references in D-Lib Magazine. *5th International Web Archiving Workshop (IWAW05), Vienna, Austria, 2005*. Available at: http://arxiv.org/ftp/cs/papers/0511/0511077.pdf

Moghaddam, A.S., Saberi, M.K. and Esmaeel, S.M. 2010. Availability and half life of web references cited in information research journals: A citation study. *International Journal of Information Science and Management*, Vol. 8, no. 2: pp. 57-75.

Moghaddam, A .I. and Saberi, M. K. 2011. The Life and death of URLs: The case of Journal of the Medical Library Association*. Library Philosophy and Practice:* Available at: http://digitalcommons.unl.edu/libphilprac/592

Moosavi, A.S., Moghaddam, A.I. and Tajeddini, O. 2012. Accessibility of online resources cited in scholarly LIS journals: A study of Emerald ISI-ranked journals. *Aslib Proceedings: New Information Perspectives*, Vol. 64, no. 2: 178-192.

Nagaraja, A., Joseph, S. A., Polen, H.H., Clauson, K.A. 2011.Disappearing act persistence and attrition of uniform resource locators (URLs) in an open access medical journal. *Program: Electronic Library and Information Systems*. Vol. 45, no. 1: pp. 99.

Nielsen, J. 1998. Fighting linkrot. *Alertbox*. Nielsen Norman Group. Available at: http://www.useit.com/ alertbox/980614.html

Reich, V. and Rosenthal, D. 2004. Preserving today's scientific record for tomorrow. *BMJ Journal*, Vol. 328, no. 7431: 61-62.

Rumsey, M. 2002. Runaway train: Problems of permanence, accessibility, and sustainability in the use of web sources in law review citations. *Law Library Journal*, Vol.94, no. 1: 27-39

Rhodes, S. 2010. Breaking down link rot: The Chesapeake Project legal information archive's examination of URL stability. *Law Library Journal*, Vol. 102, no.4: 587.

Saberi, M.K., and Abedi, H. 2012. Accessibility and decay of web citations in five open access ISI journals. *Internet Research*, Vol. 22, no. 2: 240.

Schafer, K.E., Weibel, S.L. and Jul, E. 2001. The PURL project. *Journal of Library Administration*, Vol. 34, no.1-2: 123-149.

Sellitto, C. 2004. A study of missing web-cites in scholarly articles: towards an evaluation framework. *Journal of Information Science*. Vol. 30, no.6: 484-95.

Sellitto, C. 2005.The Impact of impermanent web-located citations: A study of 123 scholarly conference publications. *Journal of the American Society for Information Science and Technology*. Vol.56, no. 7: 695–703.

Spinellis, D. 2003. The decay and failures of web references. *Communications of the ACM.* Vol. 46, no. 1: 71-77.

Tajeddini, O., Azimi, A., Sadatmoosavi, A. and Sharif-Moghaddam, H. 2011. Death of web citations: a serious alarm for authors. *Malaysian Journal of Library & Information Science*, Vol. 16, no. 3: 17-29

Tyler, D.C., and McNeil, B. 2003. Librarians and link rot: a comparative analysis with some methodological considerations. *Portal: Libraries and the Academy*. Vol. 3, no.4: 615-32.

Wagner, C., Gebremichael, M. D, Taylor, M.K. and Soltys, M. J. 2009. Disappearing act: decay of uniform resource locators in healthcare management journals. *Journal   of Medical Library Association,* Vol.97, no. 2: 129-130.

Wren, J.D. 2004. 404 not found: The stability and persistence of URLs published in MEDLINE. *Bioinformatics*. Vol. 20, no. 5: 668-672.

Wren, J.D. 2008. URL decay in MEDLINE—a 4-year follow-up study. *Bioinformatics*. Vol. 24, no. 11: 1381-1385. DOI: 10.1093/bioinformatics/btn127.