ARTICLE

The half-life of internet references cited in communication journals

DANIELA V. DIMITROVA AND MICHAEL BUGEJA
Iowa State University, USA

Abstract

This exploratory study examines the use of online citations, focusing on five leading journals in journalism and communication. It analyzes 1126 URL reference addresses in citations of articles published between 2000 and 2003. The results show that only 61 percent of the online citations remain accessible in 2004 and 39 percent do not. The content analysis also shows that .org and .gov are the most stable domains. Error messages for 'dead' URL addresses are explored. The instability of online citations raises concerns for researchers, editors and associations.

Key words

internet research • online citations • replicability • source attribution

INTRODUCTION

The internet is becoming a mainstream tool for scholarly research, chiefly because of its ability to access data quickly and conveniently. However, some studies document a troubling trend: online citations disappear at increasing rates over time (Bugeja and Dimitrova, 2005; Dellavalle at al., 2003; Germain, 2000; Taylor and Hudson, 2000; Tyler and McNeil, 2003). Indeed, such citations 'in cyberspace are like atoms in various states of decay', according to a recent *Chronicle of Higher Education* article (Carlson, 2004). The phrase 'half–life' is an

811

apt metaphor, namely, the length of time required for half of the atoms of a radioactive substance to decay. Authors employ this term to ascertain the length of time for half of a given number of internet citations to decay in a journal. By analyzing each article in a journal, the 'half-life' of internet citations in a specific publication also can be ascertained – a measurement distinct from other journal benchmarks, including the ones used by *Journal Citation Reports*.[1]

This calculation is essential in understanding an ominous phenomenon: how often do internet-based footnotes disintegrate, redirect users to other content or otherwise vanish intermittently or permanently? The footnote is the most basic element of research. Without stable footnotes, how can scholars of the future rely on internet citations? This study focused on the state of decay in online references published in the following leading journals featuring scholarship across multiple communication emphases, including new media, affiliation with influential associations and high-impact factors according to the ISI Citation Index: *Human Communication Research,* the *Journal of Broadcasting & Electronic Media,* the *Journal of Communication,* the *Journalism and Mass Communication Quarterly* and *New Media & Society*. In addition to determining the state and stability of online citations in these journals, it also sought to address the practical implications for scholars who use online citations and the broader question of whether the internet can remain a reliable tool for research.

Internet research is vital to scholarship because the medium serves as a convenient electronic warehouse of data accessible at all hours and in great quantities, thereby increasing the scope and breadth of scholarship. Much of the previous work on this erosion of internet citations has been done by library scientists. Even though some studies have recommended solutions (e.g. Evans and Furnell, 2001; Taylor and Hudson, 2000), none of them have been implemented on a cross-disciplinary scale that includes the communication journals identified and analyzed in this study. In addition, few studies have discussed the severity of the problem in the context of replicability of research. Journalism and communication departments are responsible for upholding editorial standards, especially in academic publishing, as these disciplines often educate editors or help to establish editorial standards across a wide array of content areas. Finally, as scholars and journals begin to rely more on technology – moving from paper to digital archives – it has become important to alert researchers to the issues associated with frequent use of internet references.

LITERATURE REVIEW

Web-based resources make locating relevant information fast and convenient. This is one of the reasons why the use of online sources in general (Vaughan and Shaw, 2003) and by journalism and communication scholars in particular (Chang and Tai, 2004) has increased. However, online resources are also prone to disappear from the web.

The 'ephemeral' nature of internet information has been discussed in studies in library sciences and information science, which have examined the use of online resources (Baldi, 1998; Casserly and Bird, 2003; Natriello, 1997; Tyler and McNeil, 2003; Vaughan and Shaw, 2003). Authors agree that citation accuracy and access availability of online references for future scholars are fundamental elements of reliable academic research. Among five key elements of information integrity – content, fixity, reference, provenance and context – reference is crucial: the ability to find and access sources at different points in time (Natriello, 1997). The dynamic nature of the internet compromises access to online sources available through several channels including, but not limited to, digitized journals, public and proprietary databases and search engine content.

Death rate of hyperlinks

As Tyler and McNeil (2003) note, there are two types of online information availability studies: 'diachronic' studies of random websites and 'synchronic' studies of specific online bibliographies. In one of the first diachronic studies of website longevity, Koehler (1999) established two main types of longevity behavior: constancy and permanence. Constancy refers to whether a webpage carries the same online content over time, while permanence measures the probability of a webpage to carry the same URL over time (Koehler, 1999). However, Koehler's study focused on a random sample of generic URLs. The permanence of web citations used in the leading communication journals is of interest in this study.

Koehler (1999) developed three categories for webpage permanence: always-present web documents, intermittent web documents (which fail to respond but reappear over time) and 'comatose' web documents (those vanished from the URL). About one-third of the webpages in this study could not be retrieved, leading to a half-life of 1.6 years. Only ten percent of the missing URLs were classified as intermittent webpages. In a later four-year longitudinal study, Koehler (2002) found that the half-life of webpages equaled two years. He updated the results from his earlier studies and concluded that webpage content becomes more stable over time, noting that 'the half-life of web resources in different disciplines, domains and fields differ' (Koehler, 2004: 15).

Internet citations in research

There is an increasing body of research documenting how scholars use internet sources in their work. Germain (2000) followed the accessibility of 64 URLs cited in academic journal articles. She found that the availability of online sources declined over time: 27 percent of the URLs were gone in 1997, 38 percent in 1998 and 48 percent in 1999. This profound attrition rate raises concerns for follow-up studies, which stall because of inaccessible source materials. Harter and Kim (1996) examined the availability of electronic

813

resources among peer-reviewed electronic journals. They found that even within the same year, one-third of the online citations were unavailable. Some studies report a lower attrition rate. When examining occurrences of internet references in three important US scientific journals: the *New England Journal of Medicine, JAMA: The Journal of the American Medical Association* and *Science*, researchers found that only 13 percent of internet references were not accessible after two years (Dellavalle et al., 2003). A possible explanation is that medical information online is more permanent; however, another argument might be that loss of any such information has great impact on issues involving health, wellness and scientific method. Perhaps such journal literature, accessible through value-added databases such as OVID Medline or PubMed, which are paid for by medical libraries, are more stable because of their pay-per-use content (however, this point falls beyond the scope of this article).

Taylor and Hudson (2000) studied the decay in website bibliographies and found that 30 percent of the links experienced 'linkrot'. Casserly and Bird (2003) examined 500 internet citations randomly chosen from scholarly articles published in library and information science journals. They found that only 56.4 percent of those URLs were permanent, while the rest had disappeared from the original web address. Further, the study showed that more than half of the online citations contained incomplete information and the majority did not include a retrieval date. Boynton and Imfeld (2004) found decaying online references in popular public relations textbooks; inactive links ranged from 37 percent in 1998 to 17 percent in 2001 textbooks.

Typically, a failing online citation brings up an error message in the browser. The two most common messages explaining reasons for response failure of webpages were 'server error' (i.e. no Domain Name System (DNS) entry) and 'page not found' (i.e. 404 HTML errors) (Koehler, 1999). Most unavailable URLs bring up 404 errors. Germain (2000) also found that 'server error' and 'file not found' were the two most common messages for inaccessible webpages. 'File not found' was the most frequent error message reported by Casserly and Bird (2003). Some researchers then tried to locate missing online citations through additional searches. Casserly and Bird (2003) found that close to half of the online citations they examined were initially unavailable, but increased the final result to 81.4 percent available citations by using different methods, including correcting errors in the URL, browsing the parent website or using the Google search engine.

The goal of the present study was twofold: to examine the rate of decay in online sources in journalism and communication journals and, based on that data, to recommend methods to extend the life of citations, by addressing the following research questions:

> RQ1: What is the frequency of use of internet citations in articles published in *Human Communication Research,* the *Journal Broadcasting & Electronic Media,* the *Journal*

814

of Communication, the *Journalism and Mass Communication Quarterly* and *New Media & Society*?

RQ2: What are the general characteristics of the internet citations? Specifically, how many of the internet citations are still active?

RQ3: What factors can serve as predictors of online citations' permanence?

METHOD
Sample and unit of analysis
This study focuses on refereed journal articles published in *Human Communication Research,* the *Journal of Broadcasting & Electronic Media,* the *Journal of Communication,* the *Journalism and Mass Communication Quarterly* and *New Media & Society.* These journals were chosen for analysis because they are among the most prestigious in journalism and communication; are representative of scholarship across dozens of emphases, including new media; and are affiliated with three influential associations, the International Communication Association (ICA), the Association for Education in Journalism and Mass Communication (AEJMC) and the Broadcast Education Association (BEA).[2] *New Media & Society* was selected because it is considered the leading journal focusing specifically on new media research. Another criterion used in the selection of the five journals was their impact factor provided by the ISI Citation Index.[3] All journals are published quarterly during the study period, which makes comparisons more valid. A four-year period was chosen for this exploratory study of internet citations permanence. All articles published between 2000 and 2003 in the selected journals were retrieved using the following library online databases: Expanded Academic and Proquest. This four-year period was chosen because there has been an increase in the number of internet citations being used in scholarly journals since 2000 (Davis, 2002) and it allowed for comparison of internet citation use over time.

Quantitative content analysis was conducted to capture all occurrences of internet references among all refereed journal articles published in the selected journals. Essays, book reviews and editorial notes were excluded from the analysis. Thus, the population of all journal articles was downloaded from June to October 2004. The unit of analysis was the URL.[4] Trained graduate student coders accessed all articles and recorded the URL address for each internet source cited, excluding online citations appearing as 'ibid' in the same article. The URL was then coded for a number of variables identified in prior research.

Coding process
The main variable of interest was URL permanence (Koehler, 1999), that is whether the URL cited was still active or not. This was coded as a nominal Yes/No variable. The coders accessed each online citation first by clicking on the URL address and, if this did not work, by pasting the link into the web browser.

815

This two-step process was adopted following the procedure used by Bugeja and Dimitrova (2005). Internet Explorer (IE 5.0), the most popular web browser at the time, was used by the coders. Each internet source was also coded for the following:

- top-level domain of the URL;
- error message for unavailable links;
- URL constancy, i.e. whether the online content matched the citation;
- whether the URL was hyperlinked correctly;
- retrieval date of the citation; and
- presence of a '%' sign in the URL.

Additionally, the coders deleted 'percent' signs or spaces when those were present and tested again if the link worked or not. This was important because browsers convert spaces in URLs to 'percent' signs.

The coders were instructed to revisit webpages which brought up a DNS server error because of the possibility of temporary server malfunction. After two subsequent visits, the URL was coded as dead. An intercoder reliability check was conducted on a random sample of URLs, as recommended by Riffe et al. (1998). Every eighth URL was systematically sampled and recoded. Cohen's Kappa was used to estimate the intercoder reliability between the two independent coders.[5] Intercoder reliability was established at .73 across all categories. This agreement level is acceptable (Riffe et al., 1998). High levels of agreement were achieved for several variables, including date of retrieval and top-level domain. Other categories, such as error messages and spaces in the URL, ranked lower in inter-rater agreement.

Methodological challenges

Some challenges with conducting a content analysis of online citations were observed during this study and need to be noted. Two coders accessed the online citations over a five-month period. First, some coding differences could be rooted in the difference in time since it was possible for some URLs to have disappeared. We found no instructions on how to content-analyze data that are subject to alteration and/or decay because of the manipulatable and ephemeral nature of the internet. Second, differences were observed between articles produced online in .html format versus articles produced in .pdf format. To ensure comparability between the journals, .pdf documents were examined in all cases. (Some interesting differences due to content format are discussed below.)

Finally, the five journals examined here use three different citations style guides. At the time that this study was conducted, *Human Communication Research,* the *Journal of Broadcasting & Electronic Media* and the *Journal of Communication* used the latest version of the *Publication Manual of the American Psychological Association* (APA; 1994, 2001), which requires retrieval dates for

online sources (e.g. 'retrieved month, day, year, from URL'). The *Journalism and Mass Communication Quarterly* articles were based on the *Chicago Manual of Style* (1993, 2003), which does not require retrieval dates. *New Media & Society* follows the Harvard citation style, which requires access dates for websites (e.g. 'consulted month, year'). Differences in the style guides were taken into consideration, when present, and are explained in the results section.

RESULTS
Frequency of internet sources
The first research question addressed the frequency of use of internet sources in refereed journal articles in the field of communication. A total of 1126 online citations were retrieved for the four-year period (excluding duplicates), a procedure adopted because the objective was to ascertain rather than inflate the 'half-life' phenomenon. With the exception of the 2001 publication year, the results indicate an increase in the number of online references used from 2000 to 2003. There were 201 online citations in 2000 articles, 165 in 2001, 350 in 2002 and 410 in 2003. Some articles used no online sources while one 2003 article, for example, cited 19 online sources.

The breakdown for each journal indicates that there were 50 internet sources (4.4%) in articles published in *Human Communication Research*, 278 (24.7%) in the *Journal of Broadcasting & Electronic Media*, 102 (9.1%) in the *Journal of Communication*, 264 (23.4%) in the *Journalism and Mass Communication Quarterly* and 432 (38.4%) in *New Media & Society* over the four-year period (see Table 1). The heaviest use of internet citations was observed in *New Media & Society*, a journal which focuses specifically on new media research.

Characteristics of online citations
To answer the second research question about the general characteristics of the internet citations, frequencies were run for each pertinent variable. Of the

• Table 1   Internet citations in five communication journals

JOURNAL/YEAR	2000	2001	2002	2003	TOTAL
CITATIONS WORKING	YES/NO	YES/NO	YES/NO	YES/NO	YES/NO
Human Communication Research	7/7	4/9	10/2	7/4	28/22
Journal of Broadcasting & Electronic Media	35/34	7/13	72/36	58/23	172/106
Journal of Communication	8/13	24/13	13/9	16/6	61/41
Journalism & Mass Communication Quarterly	20/18	21/21	47/37	76/24	164/100
New Media & Society	33/26	28/25	71/53	133/63	265/167
All journals	103/98	84/81	213/137	290/120	690/436

total of 1126 internet citations, 436 (38.7%) did not work at the time that this study was conducted, while the remaining 690 (61.3%) were still accessible. Of those 690 links that worked, only 408 (59%) matched the content given in the reference. Thus, only 36 percent – or 408 of the original 1126 citations – worked as promised. Some links went to the main homepage of the site as opposed to the webpage cited in the article, while others linked to webpages with completely different content. This was most likely to happen with current news events or media reports – for example, a link to Nielsen NetRatings (http://www.nielsen-netratings.com/pr7), which was no longer active.

In sum, close to 39 percent of the online citations were dead hyperlinks. This number is comparable to statistics given in some previous studies (Germain, 2000; Koehler, 1999). The number of inaccessible internet citations almost equals the number of accessible citations for one publication year: for example, 81 of the online citations from 2001 no longer worked, and only 84 were still accessible when checked in 2004 (see Table 1).

Most common were links to the .com domain, which accounted for 372 (32%) of the internet citations. Next were links from .org – 337 (30%). Only 131 (21%) of the online citations were links to .edu. There were even fewer links to the .gov and .net domains. Most of the online citations had .html (30%) or .htm (13%) extensions. There were few database-driven links – for example, .asp was present in 26 (2.3%) of the cases. Only 18 (1.6%) of the links had .pdf extensions.

Citation errors

Citation errors are an indication of the ephemeral nature of online citations. Consistent with previous studies, the most frequently cited reason for a dead hyperlink was a message that the page was not found – sometimes explicitly saying '404 error' or 'page not found'. A number of online citations required a user name and password. Subscription requirements applied to proprietary library databases as well as online news content. Here is an example from the *Los Angeles Times*: 'The story you requested is available only to registered members. Registration is free and offers great benefits. Click here to register if you are not a registered member of latimes.com'. Few of the online citations called up a redirect page. A dead link hosted by a government website, for example, brought up the following redirect message:

> As of May 31, 2002, www.nara.gov became www.archives.gov! Please update your bookmarks. Wait 10 seconds or click now to visit www.archives.gov. Thank you! (US National Archives)

The majority of the links (86 percent) did not contain an error in the URL. However, 14 percent of the citations were not hyperlinked correctly: for example, they contained spaces in the URL address, had a missing '/' sign or

incorrect file extensions. Also, it is interesting to note that the majority of the URLs did not provide a retrieval date.[6] As previously mentioned, APA (1994, 2001) and Harvard style instructs authors to provide retrieval dates. The *Chicago Manual of Style* (1993, 2003) advises authors to include access dates if the journal or discipline requires them, but also notes in the 15th edition:

> Access dates in online source citations are of limited value, since previous versions will often be unavailable to readers . . . Chicago therefore does not generally recommend including them in a published citation. (section 17:12; http://www.chicagomanualofstyle.org/ch17/ch17_toc.html)

That mixed message may explain why only 34 percent (89) of online citations contained retrieval dates in the *Journalism and Mass Communication Quarterly*, which follows the *Chicago Manual of Style*, compared with *New Media & Society*, for example, where 89 percent (383) of the Harvard style citations included retrieval dates.

Predictors of internet citation permanence

The last research question asked what factors could serve as predictors of URL permanence. Chi-square comparisons show that there were statistically significant differences across top-level domains (χ^2 (4, N = 1126) = 29.55, p < .001). The results show that .org and .gov were the most stable domains for URLs, with 70 percent and 67 percent active links, respectively.[7] The results also indicate few links to country-level domains or newer domains such as .coop.

Year of publication also was a significant predictor of URL permanence (χ^2 (3, N = 1126) = 31.48, p < .001). Unsurprisingly, the largest proportion of active links was from the 2003 journal articles, 71 percent of which still worked, compared with 61 percent from 2002, 51 percent from 2001 and 51 percent from 2000. Another determinant of whether or not the citation still worked was the website level of the citation. We expected top-level links (i.e. homepages) or second-level pages to be more stable. This was indeed the case: lower-level links were more likely to be unavailable (χ^2 (5, N = 1126) = 49.62, p < .001). For example, only 16 percent of the top-level links did not work, compared with 37 percent second-level links and 47 percent of the links at the sixth level or higher.

Half-life estimation

Following the procedure used in prior research (Koehler, 1999; Tyler and McNeil, 2003), we decided to establish a half-life for online citations in our leading journals which can be used by future scholars and journal editors. (Again, 'half-life' was defined as the time required for half of all online citations in a journal to disintegrate.) This amount of time may differ for different disciplines or different years, as noted by Koehler (1999). Our

estimates are based on the four most recent years of publication. A four-year window is reliable enough for editors to predict the rate of decay in a future publication year, enabling them to ascertain over time whether their editorial methods are extending the half-life in their journals. The following formula is used to calculate the half-life of online citations for each journal year. It assumes exponential decrease in the number of available citations. Similar assumption was made in Koehler's research (Koehler, 1999; Koehler, 2002).

$$W(t) = W(0) \; e^{at}$$

where *W(0)* is the number of working online citations at the time of publication, *W(t)* is the number of working online citations at some later time *t*, and *a* is a constant that can be calculated from the available data. Now, the half-life of internet citations, t_h, is calulated as follows:

$$t_h = [t \; ln(0.5)] \; / \; [ln \; W(t) - ln \; W(0)]$$

where t_h is the estimated number of years it takes for 50 percent of the published internet citations to stop working.

Using the second formula, we calculated the half-life for each of the five journals and each of the four publication years that we analyzed. The results are shown in Table 2. Thus, if we look at the attrition rate from 2003 to 2004 only, we can conclude that it will take 2.02 years (approximately two years) for half of the 2003 online citations to decay. The half-life tends to increase when we examine longer periods of time. This supports Koehler's (2002) finding that web documents become more stable over time. Again, half-life is contextual and may be different for different academic disciplines and different publication years.

Based on the four-year period and the five top journals examined in this study, the average half-life for journalism and communication internet references is estimated to be 3.17 years. This means that it will take about three years for half of the internet citations to vanish. Based on the figures

• Table 2   Half-life of internet citations in five communication journals

JOURNAL/YEAR	2000	2001	2002	2003	AVERAGE
Human Communication Research	4.0	1.76	7.6	1.53	3.72
Journal of Broadcasting & Electronic Media	4.08	1.98	3.41	2.08	2.89
Journal of Communication	2.87	4.8	2.63	2.17	3.12
Journalism & Mass Communication Quarterly	4.32	3.0	2.39	2.53	3.06
New Media & Society	4.77	3.26	2.49	1.79	3.08
All journals	4.01	2.96	3.70	2.02	3.17

presented in Table 1, editors can estimate the projected attrition rate for following years and note whether the half-life increases or decreases, using 3.17 as a benchmark number.

DISCUSSION

The findings of this study document substantial online citation failure and are consistent with previous research in other disciplines. Close to 40 percent of the online citations in the top five journalism and communication journals examined in this study have disappeared over a four-year period, which raises concerns for scholars who use internet citations. This problem is exacerbated because many of the online citations, while accessible, did not match the original source document, signaling even higher citation failure.

The internet, no doubt, allows researchers easy access to online resources and the ability to disseminate information quickly via the world wide web. These are substantial improvements over other eras, when researchers had to travel to government and educational agencies and spend countless hours photocopying boxes of data and files. However, universal standards to ensure accuracy and access to online citations are needed. Without the ability to fact-check citations, research may become as reliable as opinion and method as non-replicable as art. To illustrate this and the internet's dynamic but unreliable aspects, the authors realize that this study cannot be replicated because of the decay of online citations over time.

The phenomenon of decaying online citations is relatively recent. However, the diffusion of online access – including databases and digital journals – undoubtedly will create a snowball effect as more authors cite hypertext whose addresses, as this study and others have shown, may lapse. Without cogent recommendations to offset the decay of internet citations, scholars will not be able to access the full array of citations. Journals will become increasingly unreliable when future studies attempt replication, without which validity cannot be fully ascertained, all of which will contribute to the erosion of standards. The role of the academic citation is so basic to research that it is easy to overlook, even as its failure infects our best communication journals.

RECOMMENDATIONS

This study has important implications for different constituencies: educators and authors; editors and publishers; and associations and organizations. Recommendations for website bibliographers, website creators, computer administrators and governing bodies overseeing URL implementation have already been made by library and computer scientists (e.g. Evans and Furnell, 2001; Taylor and Hudson, 2000). However, the recommendations here are directed at personnel, publishers and organization officials who in previous eras would be aghast at book citations referencing wrong pages. Thus they need to

hold the internet to the same standards as the book, but also be aware of the special challenges of online references. Of particular concern is the increasing use of online citations coupled with a substantial number of citations without retrieval dates, perhaps an early (and as yet untested) warning indicating poor scholarship methods resulting from lack of universal standards. Other evidence of erosion of standards included citations that simply referenced the domain name without the particular link where specific data could be found, as in 'http://www.fcc.gov' or 'http://www.nytimes.com'.

For authors and educators

Journalism and communication scholars need to be careful when using the web as a source. We recommend that journalism and communication programs cover the half-life of internet citations at both the undergraduate and graduate levels, inasmuch as graduates of programs eventually may work in editing positions for publications using web citations and graduate students will need to understand the half-life effect in order to conduct reliable research. In the interim, they should cite journals and books rather than their online sources and/or utilize library databanks using PDF formats which, essentially, is a digital picture of the printed page. When using PDF formats, authors should provide retrieval dates to enhance validity, even when using the *Chicago Manual of Style*, which acknowledges that the half-life of online citations exists but then obfuscates the issue by leaving it up to the researcher, discipline and intended journal. If websites must be used, authors should test online citations before submitting manuscripts. When links fail, they should check formatting errors and/or update accordingly, using search engines such as Google to find the latest hyperlink to a source. Authors might be wary of citing material from websites using less reliable extensions such as .net or .int. Relatively stable extensions are .gov and .org. As a precaution, print two hard copies of web contents, including citations (one for the author's files and one for the editor's files, to be sent upon request). Authors also might want to save web documents on disk simply by using the 'File>Save As' options in a web browser or 'Print Screen' options on their desktop.[8] Current browsers also allow saving webpages as PDFs, which is recommended because it preserves well non-text content.

For editors and publishers

Our study shows that editors and publishers can do more to inform authors about the use of online citations and to extend the half-life of citations. Editors should test online citations in accepted articles in the pre-production phase of the publication process. Authors should be notified when tested links fail and asked to supply original online source documents or to use the PDF or paper version, if available. Editors also should inform authors about the proper use of, and issues associated with, use of online citations. Submissions

guidelines for journals should include examples of citing online resources. Depending on resources, they can construct searchable intranets preserving their publications/citations in their entirety.[9] Journals such as *New Media & Society* could be the leaders in the field by building an internal database with all content including online citations at the time of publication in order to preserve them for future scholars. Publishers of internet-only journals should avoid moving journal issues to new online locations since this may weaken tenure and promotion cases for scholars who have published there, due to inaccessible article links and/or inability of other scholars to cite such research.

For associations and organizations

Academic associations and other organizations representing and/or vending research have an ethical obligation to resolve problems associated with the half-life of internet citations. Specifically, they should invest in methods to archive conference and scientific presentations and associated scholarship and publications, providing stable access to online sources. Communication associations, in particular, should appoint a taskforce or ask publication supervising committees to analyze methods to address the half-life of internet citations, working with editors to call attention to the erosion of such citations and the impact on scholarship. Finally, associations and organizations should vend PDF formats of published works, as formatting issues associated with HTML text versions are complex and can undermine stability across platforms.

CONCLUSION

This study documents the rate of decay of online citations used in leading communication journals. It also provides recommendations to address the half-life phenomenon. Unless authors, educators, editors and associations take dramatic steps to extend the half-life of internet citations, replication may prove difficult in the social, medical and natural sciences, precisely because the footnote remains the most basic component of research without which everything else is tainted, including research methodologies and scientific formulas. While it is true that medical citations are, in general, more stable than communication citations, even the decay of one can have a negative impact on scholarship.

Specifically at risk is research about the internet itself, from the effects of mediated communication to the impact of information technology policy, as such scholarship by its nature must reference online resources. The study raises broad questions not only associated with technology but also media history. Because the dynamic nature of the world wide web offers initial ease of access but unreliable subsequent retrieval, and with decay rates increasing over time, the internet may prove to be an inhospitable medium, especially

for web–based research. This particular contention, to be explored in future studies, takes into account a factor which has yet to be investigated fully in the literature – namely, the possible end of the printing press era (during which there were paper transcripts, even for television) and the beginning of the internet one (which no longer relies on paper transcripts). These are two opposite media, with one ensuring exact replicates of originals and the other allowing users to manipulate those originals. That feature – the ability to manipulate – is at odds with the stability that footnotes have required in accountable scholarship.

Acknowledgement

Notes
1 The 'internet citation half-life', which will vary from journal to journal based on the number of online citations in the publication's annual index of articles, is distinct from the terms 'cited half-life' and 'citing half-life' used by Journal Citation Reports, benchmarking the age of received and referenced articles, respectively. The cited half-life determines when about half of citations are under and half over that calculation (see Journal Citation Reports, 2004). The term 'half-life' was used in a similar fashion to the current research by Tyler and McNeil (2003), who estimated the half-life of *College & Research Libraries News* web bibliographies to be between four and five years over an eight–year period of study.
2 Oxford University Press publishes *Human Communication Research* and the *Journal of Communication* on behalf of the International Communication Association. The *Journalism and Mass Communication Quarterly* is published by the Association for Education in Journalism and Mass Communication. Broadcast Education Association publishes the *Journal of Broadcasting & Electronic Media* through Lawrence Erlbaum Associates. *New Media & Society* is published by Sage.
3 The ISI Citation Index calculates impact factors for scholarly journals. Based on its 2003 ranking, *Human Communication Research* has the highest impact factor of all communication journals, 1.612. The impact factor for the *Journal of Broadcasting & Electronic Media* is 0.44, the *Journal of Communication* is 0.793, the *Journalism and Mass Communication Quarterly* is 0.393 and *New Media & Society* is 0.689.
4 URL stands for Uniform Resource Locator and refers to the web address of the hyperlink. Typically, it begins with 'http://', for example, http://www.nytimes.com is the URL for the *New York Times*.
5 Cohen's Kappa was used to estimate the intercoder reliability, using the following formula: $Kappa = (P_0 - P_e)/(1 - P_e)$, where P_0 is the observed agreement between the coders and P_e is the expected agreement based on chance.
6 The oldest retrieval date given in our sample was July 1996 (without an exact day) and was cited in a 2000 article in the *Journal of Communication* (http://www.census.gov/population/www/estimates/uspop.html). The link did not work at the time that this study was conducted.
7 Future studies should incorporate hypotheses about URL characteristics and URL permanence based on the significant predictors identified in this study.

8   While there are some internet archives in existence – for example, *The Wayback Machine*, which has collected large portions of the internet since 1996 – none of these archives is capable of recording the complete world wide web content.

9   One such example from the hard sciences is *CiteSeer* (http://citeseer.ist.psu.edu/). *CiteSeer* is a scientific literature digital library that interlinks all science research, ranks active bibliography entries, dynamically stores the data at a Penn State server and even graphs the year of publication of citing articles.

References

Baldi, S. (1998) 'Normative Versus Social Constructivist Processes in the Allocation of Citations: A Network-analytic Model', *American Sociological Review* 63(6): 829–46.

Boynton, L.A. and C. Imfeld (2004) 'Virtual Issues in Traditional Texts: How Introductory Public Relations Textbooks Address Internet Technology Issues', *Journalism and Mass Communication Educator* 58(4): 330–42.

Bugeja, M. and D.V. Dimitrova (2005) 'Exploring the Half-life of Internet Footnotes', *Iowa Journal of Communication* 37(1): 77–86.

Carlson, S. (2004) 'Here Today, Gone Tomorrow. Studying How Online Footnotes Vanish', *Chronicle of Higher Education*, 30 April, URL (consulted 3 June 2007): http://chronicle.com

Casserly, M.F. and J.E. Bird (2003) 'Web Citation Availability: Analysis and Implications for Scholarship', *College and Research Libraries*: 64(4): 300–17.

Chang, T.K. and Z. Tai (2004) 'Mass Communication Research and Invisible College Revisited: A Field with a Few Ideas, but No Classics?', paper presented at the Theory and Methodology Division, Association for Education in Journalism and Mass Communication, Toronto, Canada, 4–7 August.

Chicago Manual of Style (14th edn) (1993) Chicago, IL: University of Chicago Press.

Chicago Manual of Style (15th edn) (2003) Chicago, IL: University of Chicago Press.

Davis, P.M. (2002) 'The Effect of the Web on Undergraduate Citation Behavior: A 2000 Update', *College and Research Libraries* 63(1): 53–60.

Dellavalle, R.P., A. Drake, M. Graber, L. Heilig, E. Hester, J. Kuntzman and L. Schilling (2003) 'Going, Going, Gone: Lost Internet References', *Science* 302(5646): 787–88.

Evans, M.P. and S.M. Furnell (2001) 'The Resource Locator Service: Fixing a Flaw in the Web', *Computer Networks* 37(3–4): 307–30.

Germain, C.A. (2000) 'URLs: Uniform Resource Locators or Unreliable Resource Locators', *College and Research Libraries* 61(4): 359–65.

Harter, S.P. and H.J. Kim (1996) 'Electronic Journals and Scholarly Communication: A Citation and Reference Study', *Information Research* 2(1), URL (consulted 3 June 2007): http://InformationR.net/ir/2-1/paper9a.html

Journal Citation Reports (2004) 'Factsheet', URL (consulted 2 July 2004): http://www.isinet.com/media/presentrep/facts/0108jcrfs.pdf

Koehler, W. (1999) 'An Analysis of Webpage and Website Constancy and Permanence', *Journal of the American Society for Information Science* 50(2): 162–80.

Koehler, W. (2002) 'Webpage Change and Persistence – A Four-year Longitudinal Study', *Journal of the American Society for Information Science and Technology* 53(2): 162–71.

Koehler, W. (2004) 'A Longitudinal Study of Webpages Continued: A Report after Six Years', *Information Research* 9(2), URL (consulted February 2004): http://InformationR.net/ir/9–2/paper174.html

Natriello, G. (1997) 'Attribution and Access: Citing Electronic Sources', *Teachers College Record* 98(3): 373–80.

Publication Manual of the American Psychological Association (4th edn) (1994) Washington, DC: American Psychological Association.

Publication Manual of the American Psychological Association (5th edn) (2001) Washington, DC: American Psychological Association.

Riffe, D., S. Lacy and F.G. Fico (1998) *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Taylor, M.K. and D. Hudson (2000) ' "Linkrot" and the Usefulness of Website Bibliographies', *Reference and User Services Quarterly* 39(3): 273–80.

Tyler, D.C. and B. McNeil (2003) 'Librarians and Link Rot: A Comparative Analysis with Some Methodological Considerations', *Portal: Libraries and the Academy* 3(4): 615–32.

Vaughan, L. and D. Shaw (2003) 'Bibliographic and Web Citations: What Is the Difference?', *Journal of the American Society for Information Science and Technology* 54(14): 1313–22.

DANIELA V. DIMITROVA is an assistant professor in the Greenlee School of Journalism and Communication, Iowa State University. She has been researching the half life of internet footnotes for several years. She is a widely published author and co-author in scholarly journals, including the *Newspaper Research Journal,* the *Harvard International Journal of Press and Politics, Public Relations Review,* the *Journal of Computer-Mediated Communication, Journalism Studies, Telecommunication Policy* and other peer-reviewed publications.
Address: Greenlee School of Journalism and Communication, Iowa State University, Ames, IA 50011, USA. [email: Danielad@iastate.edu]

MICHAEL BUGEJA is Professor and Director of the Greenlee School of Journalism and Communication, Iowa State University. He has been researching the half life of internet footnotes for several years. He is a widely published ethicist with articles in the *Journalism Quarterly, Journalism Education, Mass Media Ethics* and the *Journal of Mass Media Ethics*. Recently he has published two books: *Interpersonal Divide: The Search for Community in a Technical Age* (Oxford University Press, 2005) and *Living Ethics: Across Media Platforms* (Oxford University Press, 2007).
Address: Greenlee School of Journalism and Communication Iowa State University, Ames, IA 50011, USA. [email: Bugeja@iastate.edu]