THE PUTREFACTION OF DIGITAL SCHOLARSHIP:

HOW LINK ROT IMPACTS THE INTEGRITY OF SCHOLARLY PUBLISHING

By

MARSHAL A. MILLER

A doctoral dissertation submitted to the
College of Education
in partial fulfillment of the requirements
for the degree Doctor of Education
in Organizational Leadership

Southeastern University
November 14, 2022

THE PUTREFACTION OF DIGITAL SCHOLARSHIP:

HOW LINK ROT IMPACTS THE INTEGRITY OF SCHOLARLY PUBLISHING

by

MARSHAL A. MILLER

Dissertation Approved:

Grace Veach

Grace Veach, PhD, Dissertation Chair

Thomas Gollery

Thomas J. Gollery, PhD, Committee Member

Laura-Edythe Coleman

Laura-Edythe S. Coleman, PhD, Committee Member

Lisa A. Coscia

Lisa A. Coscia, EdD, Dean, College of Education

DEDICATION

For Robert Nelson McIntyre (Poppop) and Margaret June McIntyre (Mommom), without whom I would have accomplished nothing.

ACKNOWLEDGMENTS

There are a number of people who were instrumental in guiding me to this milestone. First, I'd like to thank my family for providing me with the opportunity to acquire a good education and the resources to tinker on my own. I thank Terri Miquel for encouraging my love of technology at an early age. Credit goes to Scott Questad for demonstrating that you can learn both in and outside the classroom. Katelyn McKinley has always been my number-one fan. A great debt is due to Cierra and Tim Robbins for encouraging me and pushing me forward.

I would also like to recognize T. Philip Nichols for familiarizing me with new ideas and teaching me how to engage them. I owe my interest and influence on this subject to the research and works of Martin Klein, Lawrence Lessig, Guido van Rossum, Linus Torvalds, and Jonathan Zittrain.

Lastly, I thank the members of my committee. First, I thank my third reader, Dr. Laura-Edythe Coleman, for her time as my professor and for the feedback she provided during the process. Next, great appreciation goes to Dr. Thomas Gollery, who served as my methodologist. His encouragement and enthusiasm for my study helped push me toward the finish line. Finally, I thank my chairperson, Dr. Grace Veach, who had to endure my dawdling pace and poor grammar. I am grateful to all three of my committee members for their investment in my dissertation.

Abstract

Research sits at the core of scholarship. The integrity of that research allows fields of study to grow and build upon one another to form the foundation for and extension of human knowledge. In the last 10 years, a new phenomenon has occurred as digital scholarship has become more prolific. This phenomenon is called link rot. Link rot occurs when over time, digital resources become inaccessible because their originally cited location has been relocated or become permanently unavailable. This study examined the extent to which link rot has affected scholarly research and how it might affect the future of digital scholarship. Historical archived data were compiled and analyzed using a self-created tool to evaluate the extent to which a publication has been affected by the phenomenon of link rot. Study data were accessed through content analysis of 2,500 published, peer-reviewed scholarly articles, representing a span of 10 years (2013–2022) of data collection. Five specific academic domains in the scholarly literature were identified for study purposes: (a) arts and humanities; (b) business; (c) health and medicine; (d) science, math, and technologies; and (e) social sciences. The study showed that 36% of all links were broken, and 37% of digital object identifiers were broken. The study also showed a significant difference in the percentage of broken links between academic disciplines, as well as the percentage of broken digital object identifier links.

Keywords: Link rot, scholarly publishing, citations, DOI, persistent identifiers, links, references, research, research policy.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

I. INTRODUCTION

Research represents the core of scholarship, and the integrity of that research allows fields of study to grow as scholars build upon one another's work to form the foundation and extension of human knowledge (West et al., 2017). In the last 10 years, a new phenomenon has occurred as digital scholarship has become more prolific (Shemberko & Shershova, 2018). This phenomenon is called link rot. *Link rot* occurs when, over time, digital resources become inaccessible because their originally cited location is relocated or becomes permanently unavailable. This study examined the extent to which link rot has affected scholarly research and how it might affect the future of digital scholarship.

Background of the Study

The internet has become so ingrained in everyday life that its true age can seem indeterminable. Many early iterations of the internet existed, but the first webpage was not published until December 1990. Much like a human life, the internet required underwent a process of maturation before it became the entity that it is today. An example of the internet's early lack of sophistication can be seen at the turn of the millennium when users feared the internet would crash because of the clock and calendar features. Today users know much more about the internet and how to maintain it, but many of its foundational protocols have aged poorly and have begun to cause difficulties that can be observed over a period of time (Krol & Zdonek, 2020). Link rot represents one such issue (Klein & Balakireva, 2021).

The first web page published in December of 1990 is no longer accessible, and an increasing number of resources are following its fate. The internet did not serve as a reference resource until much later in its life. It took some time for researchers to recognize the degradation of internet resources that was occurring as technology advanced and individuals in society became increasingly dependent on those resources (Krol & Zdonek, 2020).

Once the internet became the primary way of accessing resources for individuals in the majority of the world, governments and organizations began to consolidate their offerings to internet repositories (Shemberko & Shershova, 2018). The United States took part in this initiative. In 2011, a few researchers at Harvard Law School, under the umbrella of the Berkman Center for Internet and Society, began to notice that some of the references and links on the Supreme Court's website had become inaccessible, and the website offered no alternative way to locate them (Zittrain et al., 2013). Given the essential nature of Supreme Court rulings, the inaccessibility of these documents presented a major flaw in their system.

Zittrain et al. (2013), a group of researchers working at the Berkman Center for Internet and Society, completed and published their study of this phenomenon in 2013. The authors found that over half of the resources cited on the Supreme Court's website were "dead" or no longer accessible. The team coined the term "reference rot" to refer to the slow degradation of references on websites over time.

Studies in the field of law have continued since the publication of the study from The Berkman Center for Internet and Society (Perkel, 2015). The gradual increasing inaccessibility of links on the Internet began to spread widely, far from being contained to the field of law (Urhan et al., 2018). As a result, the colloquial term was amended to "link rot" because the bulk of links on the Internet are not used as references but rather as a navigation tool (Jackson, 2013).

The information science community has made efforts to address this issue with the implementation of persistent identifiers (Robinson et al., 2018). The International DOI Foundation (2020) called this an open researcher and contributor identifier (ORCID) and explained:

> An identifier is a label which gives a unique name to an entity: a person, place, or thing. Unlike URLs, which may break, a persistent identifier reliably points to a digital entity. An ORCID is an example of a persistent identifier for a person.

Digital object identifiers (DOIs) and ArXiv, pronounced archive, represent two of the most popular persistent identifiers used right now (Scholarly Communication and Open Publishing, 2020). Although these solutions exist to combat the issue of link rot, some research has shown that the use of persistent identifiers is plagued by many of the same concerns as traditional links (Klein & Balakireva, 2021).

A pilot study was conducted using publications from Southeastern University's internal scholarly research repository, FireScholars, to determine the number of references that are no longer accessible in these publications. These publications included 41 dissertations from the college of education. Of the 2,090 links referenced in these publications, 887 were no longer accessible from their cited source (see Appendix). While the majority of dissertations contained at least one DOI, 76% of these DOIs were either improperly cited or deemed inaccessible.

Conceptual Framework

The framework for this study was conceptual. From a broad view, the process involved finding a PDF version of the publication, running the script developed for this study, and reviewing the report for results. The script created a list of all URLs cited in the PDF; then it attempted to navigate to each URL. The three major questions the script ran through were:

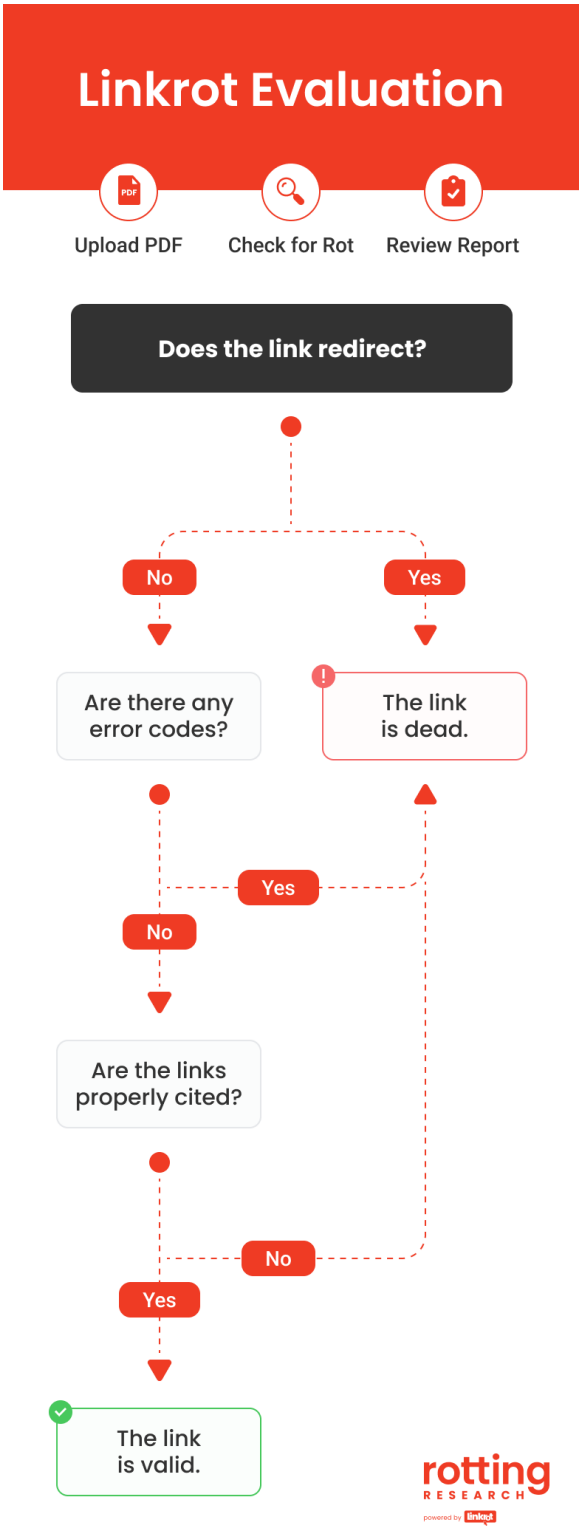
1. Does the link redirect somewhere else?

2. Does the link return any error codes?
3. Is the link properly cited in accordance with best practices and the appropriate style guide?

Figure 1 illustrates the conceptual framework's process.

Figure 1

Conceptual Framework Process



Problem Statement

West et al. (2017) asserted that link rot has severe consequences for academics who rely heavily on previous research. The authors added replication is difficult, if not impossible, without the ability of researchers to find needed resources. The loss of research citations also calls into question the validity of literature reviews in scholarly publications (Rubbo et al., 2019). Although researchers have investigated this topic in individual disciplines, few have compared the effects of link rot across disciplines (Aksnes et al., 2019). The researcher in this study conducted a pilot test of education publications and produced comparable results to those of Zittrain et al. (2013).

Purpose Statement

The purpose of this descriptive study was to measure, report, and analyze the effect of link rot on scholarly research with the intention of informing university policy on scholarly publishing and the integrity of institutional research.

Significance of the Study

This study is significant because it investigated a potential deficiency infiltrating scholarly publishing and the research it produces. The findings of this research showed the problem to be prolific. The study helps inform policy makers in higher education and scholarly publishing on ways to prevent the issue from spreading and to better preserve academia's knowledge base for the future.

Overview of Methodology

Chapter 3 provides a complete description and analysis of the methodology. Historical archived data was compiled and analyzed using a self-created tool to evaluate the extent to which a publication has been affected by link rot. Twenty-five hundred published, peer-reviewed scholarly articles were collected for the purpose of this study. Publications from between 2013

and 2022 were selected from a variety of journals and other scholarly publishing outlets representing one of five subject matters: (a) arts and humanities; (b) business; (c) health and medicine; (d) science, math, and technologies; and (e) social sciences. Each subject matter was represented by 500 articles. The sample also included 250 publications from each year in the selected range.

A command-line script was run on all publications. The script mined the publications for references to online resources. These references were tested for connection to their original source. The response to the connection request was recorded. The request either returned successfully or with an error code. All codes were recorded and organized by code. Due to the nature of this study, the tool also specifically mentioned any links directed to a DOI or Arxiv source. The source code for the tool can be found at https://github.com/rottingresearch/linkrot.

The researcher set parameters in order to avoid any bias or inconsistent judgment. As such, the following guidelines were established to standardize the results. Email addresses were noted but not checked for active status due to the spam risk. Redirects were treated as broken links. Resources requiring credentials to view an abstract or details were regarded as broken links. No navigation past the initial URL entry was allowed for verification. Every error was validated with an examination of every link returned with an error code to check for accuracy.

All publications and resources were gathered from archived and historical data. Although confidentiality presented no concern, measures were taken to remove any possibility of bias. All articles were categorized into their specified subject area and randomly numbered 1–500 with the year they were published. The output of the command-line script contained no identifying data. This output only contained a list of all links mined from the publication, the type of resources, successful link requests, and failed requests with their respective error codes.

Analyzed metrics included percentage of publications affected, average number of links broken, and breakdown of errors by type. The results were also broken down by publication year and subject matter.

Research Questions

The study addressed the following research questions:

1. To what degree is link rot evident in scholarly research?
2. To what degree is link rot evident for DOIs?
3. Considering the five areas of research identified for study purposes, in which of the error domains is the greatest degree of error manifested?
4. Considering the five areas of research identified for study purposes, in which area is the greatest degree of link rot evident in scholarly research?

Research Hypotheses

1. To what degree is link rot evident in scholarly research?

   H_0: There will be no statistically significant degree of link rot evident in scholarly research.

2. To what degree is link rot evident for DOIs?

   H_a: There will be no statistically significant degree of link rot for DOIs evident in scholarly research.

3. Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for link rot?

   H_a: There will be a statistically significant effect for academic domain on the percentage of broken links.

4. Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for DOI link rot?

H_a: There will be a statistically significant effect for academic domain on the percentage of broken DOI links.

Overview of the Analysis

Preliminary segue analyses were conducted in advance of the formal data analysis associated with the study's research questions. Evaluations of missing data, internal reliability, demographic identifying information, and descriptive information associated with the study's dependent variables were conducted using descriptive and inferential statistical techniques. The analysis, interpretation, and reporting of study findings were conducted using the 28th version of IBM's Statistical Package for the Social Sciences (SPSS).

Research Question 1 required descriptive statistical techniques that included frequencies, percentages, mean scores, and standard deviations. The nonparametric binomial test was used for statistical significance testing purposes. The test proportion value was set at the .50 level. The probability level of $p \leq .05$ represented the threshold value for statistical significance. Research Question 2 required the use of frequencies, percentages, mean scores, and standard deviations. The nonparametric one-sample chi-square test was used for statistical significance testing purposes. The probability level of $p \leq .05$ represented the threshold value for statistical significance. Research Question 3 required frequencies, percentages, mean scores, and standard deviations for comparative purposes. The one-sample *t* test was used to assess the statistical significance of the observed link against a predetermined null value representing an inconsequential threshold value for link rot in relation to literature review robustness.

In Research Question 4, frequencies, percentages, mean scores, and standard deviations represented the descriptive statistical techniques used. The one-sample *t* test was used to assess the statistical significance of the observed link against a predetermined null value representing an inconsequential threshold value for link rot in relation to literature review robustness. Cohen's *d*

was used to assess the magnitude of effect for finding in Research Question 4. Sawilowsky's (2009) conventions for effect size interpretation were used to translate the numeric effect size value into a qualitative descriptor (i.e., small, medium, large, very large, and huge).

Limitations

The limitations of this study centered around the availability of resources and the capacity to compute and codify results. The primary limitation of this study involved the availability of an optimal sample to represent the current state of scholarly publishing. The researcher had access to a limited number of scholarly publications. The limited accessibility of scholarly publications hindered the ability to fully extend the sample to perfectly represent reality. The inability to exactly represent the current state of scholarly publishing may have resulted in some biases and abnormalities in the results. The accessibility of scholarly publications restricted this study and any future studies.

The second limitation of this study involved the researcher's capacity to compute and codify the results. The researcher could only analyze a limited sample size due to the computing power and time required to process a larger number of publications. Each publication required processing, codification, and review for errors or anomalies.

Definitions of Key Terms

The following words and phrases were key terms for the study.

- **link rot**: the phenomenon of resources becoming inaccessible over time when their originally cited location is relocated or becomes permanently unavailable.
- **command-line**: a computer interface that allows tasks and processes to be executed via text strings.
- **script**: a sequence of computer code strung together to run a unified process or task.
- **mining**: the process of searching through data to find a resource.

- **persistent identifier**: a long-lasting link to a resource that remains active despite any change in data structures.
- **publication**: a work that is made accessible to the public.
- **robust**: strong, rigorous, and steadfast.
- **DOI**: The DOI system provides an infrastructure for persistent unique identification of objects of any type.
- **arXiv**: arXiv is a curated research-sharing platform open to anyone.
- **broken link**: a link that returns an HTTP status code other than a successful 200-range code.
- **content drift**: the process of resources changing form or content over time or since the initial citation date.

Summary

Research provides the core of scholarship. The integrity of research allows fields of study to grow and build upon one another to form the foundation and extension of human knowledge (West et al., 2017). The literature review resides at the center of the growth of research (Shemberko & Shershova, 2018). Literature reviews allow readers to see the principles undergirding a given publication and enable future researchers to replicate or authenticate a study (Broadus, 1983).

In the last 10 years, a new phenomenon has occurred as digital scholarship has become more prolific. This phenomenon is called link rot. Link rot occurs when digital resources cease to be accessible over time because their originally cited location has been relocated or has become permanently unavailable. This study examined the extent to which link rot has compromised past scholarly research by degrading literature reviews and how it might jeopardize the future of digital scholarship.

II. REVIEW OF LITERATURE

Link Rot

Krol and Zdonek (2020) asserted: “Link rot is the process by which website links become irrelevant or broken over time, because the actual websites they link to disappear, change content, or move to a new location”. The authors have pointed out that many scholars have referred to current times as the “digital dark ages” and explained the term was coined because of the vast amounts of digital information already lost forever. The internet has proven a volatile way to store information (Robinson et al., 2018).

Krol and Zdonek (2020) explained the *digital dark ages* refers to a system where industries from healthcare, finance, or any business connected to the internet experience peril in the hands of an unreliable internet. The authors attributed this trend in part to ever-changing technology. Another potential cause of the digital dark ages involves the increased amount of content and the fading reliability of information stored on the internet (Robinson et al., 2018). Krol and Zdonek asserted that digital information in varied formats and by different carriers can degrade if not properly stored. The authors pointed out that link rot has become so widespread that the public has become accustomed to it, and they cited a looming concern that the internet will cease to be interconnected and become a set of isolated information silos.

Scholarly articles are not immune to this phenomenon (Klein & Balakireva, 2021). Krol and Zdonek (2020) explained that the growing popularity of the internet has made it impossible to avoid citing digital resources. This issue includes references to scientific articles and scholarly

journals, previously a staple of the printed medium. The problem affects a variety of disciplines; in fact, researchers have found it exists in most, if not all, disciplines (Bradshaw et al., 2021). Krol and Zdonek claimed over 70% of internet addresses cited in articles published in the *Harvard Law Review* failed to navigate to the original content. The authors noted that even the Supreme Court has been affected, with only about 50% of links in their decisions leading to the content intended.

Citations tend to decay over time because of a preference for recentness, even over seminal works (Gerow et al., 2018). Krol and Zdonek (2020) found that only 30% of links in the sciences functioned after 4 years. This is especially troublesome when considering the age of some citations that have changed a field. In 1933, Fritz Zwicky published a comprehensive study that provided evidence about the size of galaxy clusters (Abt, 2018). Abt (2018) claimed no one but Zwicky cited the study for 27 years and added that it was not seriously discussed for a total of 41 years after its publication. Abt pointed out that scholars now see it as a seminal work with over 1,030 citations. Citations symbolize the prolongation of the a scholarly publication's lifespan (Hassan & Serenko, 2019). With the average lifespan of articles in this field hovering around 5 years, Zwicky's article would likely have been lost if it were published today, setting the research on the size of galaxy clusters behind by decades (Krol & Zdonek, 2020).

Krol and Zdonek (2020) asserted that it is essential to remember that files, formats, and software all eventually become obsolete. The authors explained that data currently available in digital form may not be readable by future devices or software. According to Krol and Zdonek, a review of the Web of Science citation index showed that the average lifespan for a resource's website was 9.3 years. Unfortunately, only 62% of that data was archived. The authors concluded this statistic reflects poorly on the resource itself, perhaps even signifying its irrelevance.

Scholarly Publishing

Isaac Newton is often attributed with saying, “If I have seen further, it is by standing on the shoulders of giants.” Hassan and Serenko (2019) asserted this aphorism might best represent the foundation of scholarship and scholarly research. They explained that this link between past and present research makes the progression of knowledge and science possible. Hassan and Serenko explained that the purpose of research is to “contribute toward the growth of knowledge and, ultimately, promote scientific progress” (Hassan & Serenko, 2019). The authors claimed that by reviewing and verifying previous research, scholars can establish a solid foundation upon which to build.

During the unwieldy and tumultuous research process, “authors create, share, and promote new concepts, theories, methods, data, and findings” (Gerow et al., 2018). Some research may contribute to seminal scientific theories, and other works might fill in the details on the sum of the research on a topic (Aksnes et al., 2019). The importance of scholarship in academia cannot be overlooked. Scholars rely on the ability to publish their research in order to establish their standing in the academic community (Jones, 2021). The production of scholarship affects an individual’s rank as a professor, eligibility for tenure, funding, and recognition (Jones, 2021). Additionally, an institution’s research output contributes to its rating and placement among top-tier schools (Zhao & Qiao, 2017). This scholarly research informs future research, influences policy decisions, and improves professional procedures (West et al., 2017).

The inclusion of references in scholarship reveals the foundation on which a researcher bases their research questions and hypotheses and enables them to differentiate their work from that of others (Penders, 2018). Researchers cite other works to indicate what principles they have drawn upon and to show how they have adopted an existing lexicon and situated their ideas among other ideas, theories, methodologies, or perceptions without explicit mention (Gerow et

al., 2018). Researchers only cite a tiny portion of an article's knowledge base because it would be impossible to cite every piece of knowledge offered on a topic in the previous literature informing a new study (Aksnes et al., 2019). References do, however, provide a method for others to retrace the steps a researcher took to establish their position (Penders, 2018).

Citations

Citations reside at the core of scholarly work (Lerner & Oddis, 2017). They can refer to anything, from a journal article to a website (West et al., 2017). Researchers typically view citations as the most obvious and essential piece of an academic text (Kafes, 2017). Researchers understand their obligation to use citations to support their ideas, seeing it as more than simply good practice (Kafes, 2017). Researchers have gone as far as to call it a moral (Penders, 2018) and ethical issue (Rubbo et al., 2019).

Referencing acts, at its core, as a method of persuasion (Aksnes et al., 2019) that some call an important feature of the language of academia (Jomaa & Bidin, 2019). Aksnes et al. (2019) explained that authors use citations to convince the scholarly community of the value and importance of their publications. Thus, they added, authors use references as rhetorical tools. Citations can also be used to indicate gaps in previous studies. This method can help researchers position themselves within previous research (Jomaa & Bidin, 2019). Additionally, citations legitimize ideas, solidify theories, and establish claims as fact (Penders, 2018).

Researchers construct knowledge by building on previous studies, and citations play a critical role in that process (Hassan & Serenko, 2019). To engage with this process, researchers are obliged to acknowledge contributions from others, including the works they draw upon (Aksnes et al., 2019). Researchers also cite sources to substantiate their findings (Rivkin, 2020). Thus, attribution leaves a trail that connects the building blocks of research to one another, linking ideas to contextualize research and support new claims (Kafes, 2017). The use of

citations provides a means for creating that trail (West et al., 2017). This process creates "a web of knowledge that carries meaning and allows other researchers to identify work as relevant" (Penders, 2018).

The use of citations can also help readers distinguish between an amateur academic and a seasoned one (Kafes, 2017). Expert and novice writers in the same academic community can vary in their citation use (Jomaa & Bidin, 2019). Kafes (2017) explained novice writers are limited in their use and range of citations, whereas an expert strategically uses citations to substantiate their claims and provide support. Jomaa and Bidin (2019) associated the varying use of citations with achieving academic success. Kafes claimed novices use citations strictly for attribution and added that experts are more likely to use citations for purposes beyond attribution. In the latter sense, they use citations to contextualize their work amid various schools of thought and among past studies, as well as to set up their arguments.

Citations signal the author's point of view, allegiance, authority, and persuasiveness (Kafes, 2017), but in academia, citations serve a far greater purpose than just providing context. Gerow et al. (2018) argued that citations offer an incomplete representation of influence. The authors added that there are many reasons why an author may cite a work, which means a reference list represents a complex meld of purposes. Small (1982) identified five distinct reasons for citation as cited in Aksnes et al., (2019). He asserted that a citation could be used to "be (a) refuted, (b) noted only, (c) reviewed, (d) applied, or (e) supported by the citing work" (Aksnes et al., 2019).

Citations also represent a form of academic currency (Penders, 2018) because they offer one of the only ways an academic writer can claim their existence in the scholarly community (Kafes, 2017). Thus, scholars use citations to declare allegiance to a community, show scholarly alliance, or indicate ambitions (Gerow et al., 2018). While often taken for granted, citations can

provide a metric of impact, which is a cornerstone of quality (Aksnes et al., 2019). Researchers commonly reference authors who have a favorable reputation in the community (Hassan & Serenko, 2019). Aksnes et al. (2019) explained that the process of using citations to measure the scholarly impact is generally referred to as citation analysis.

Methods of citation vary from one discipline to another (West et al., 2017). This occurs most visibly in a discipline's preferred citation style guide. Although each institution has different standards, most assign a certain style guide to a specific area of study (American University Librarians, 2022). American University uses the *Publication Manual of the American Psychological Association* (APA) for subjects such as business, criminology, education, international studies, linguistics, and psychology (American University Librarians, 2022). The American University Librarians (2022) explained that the fields of communications, literature, philosophy, and religion use the handbook of the Modern Language Association (MLA), and the fields of anthropology, art, and computer science utilize *The Chicago Manual of Style*. Although these three guides represent the most well-known style guides, the authors added the fields of history and music use Kate L. Turabian's *A Manual for Writers of Research Papers, Theses, and Dissertations*, and journalists use *The Associated Press Stylebook*, which is the industry standard for professional journalist. Although these guides represent examples from one institution, they paint a picture of how different areas of study conform to different citation styles, resulting in nonuniform citations across disciplines, even when citations are correctly formatted.

Unfortunately, some newer developments in information transfer have compounded existing issues with citations and link rot. As the internet becomes more established in academia, scholars must rely on digital content vulnerable to link rot with no analog equivalent (Krol & Zdonek, 2020). Kaufman and Campana (2019) identified another trend that may increase the effects of link rot on citations, pointing to the prevalence of open education resources. The

authors explained open education resources often link to large amounts of outside web materials, including videos, images, podcasts, and lesson plans. Users can only access those resources via links that can degrade and leave the artifact inaccessible and, thus, irrelevant after a few years. The authors pointed out that a link leading to a lesson's core concept can undermine the entire pedagogical plan. Every reference promises a source of information, and each one that goes unfulfilled may lead to a loss of trust and reliability (Krol & Zdonek, 2020).

Citation Errors

While the degradation of links over time functions as a major contributor to link rot, a surprising number of links fail to function properly because the researcher cited them improperly. These errors range from minor to those that alter the original intent of the cited source (Rivkin, 2020). In the field of knowledge management, researchers have found that 30% of all citations present problems (Hassan & Serenko, 2019). Rivkin (2020) showed the citation error rate per discipline ranges from 25 to 54%. This issue even affects quotations. In two different studies of medical journals, de Lacey and Mogull found a misquotation rate of 15% (Rivkin, 2020).

Citation errors and misquotations have serious implications for the integrity of published research. If a link fails, it prevents readers from finding the referenced materials, and the article's reliability comes into question (Krol & Zdonek, 2020). Trust in the reliability of published material is essential to the function of scholarly publications and the reputations of their authors (Rivkin, 2020). Notably, "scientific work does not end with publication" (Rubbo et al., 2019), so researchers must understand that misquotations can be repeated in subsequent publications and may turn into accepted knowledge (Rivkin, 2020). In scholarship, "if there are issues with one part of the data, the entire article is in doubt" (Rubbo et al., 2019).

An errant citation can also create issues for the replication and peer review processes that undergird all scholarship (West et al., 2017). An important aspect of academic work involves having a work scrutinized and peer-reviewed by members of the academic community (Rubbo et al., 2019). If readers cannot confirm a citation or examine a source, they cannot complete the verification process (West et al., 2017).

All academicians must take responsibility for citing well and responsibly (Penders, 2018). Rubbo et al. (2019) asserted that all involved in the process of establishing knowledge bear accountability, not just the authors. For example, some responsibility lies with researchers, reviewers, and editors. In his 2020 study, Rivkin found two to five "seriously misleading" citations in each of 50 randomly selected articles from six highly regarded journals.

Students learn proper methods of citation and how to choose suitable sources, but citation errors occur equally among articles published by students and those published by the most prestigious of scholars (Rivkin, 2020). Genzinger and Wills (2017) asserted librarians are typically tasked with educating students about the citation process. The scope of their work involves helping students find relevant resources, but it also involves explaining the citation process and its value. The authors explained librarians typically assume that references are accurate as long as they match the style guidelines. This, however, does not ensure that the citations are free of errors or that the meaning of the referenced material is accurately portrayed.

To dig deeper into this issue, Genzinger and Wills (2017) looked at three of the most prominent scholarly journals in library science. The authors wondered how those tasked with instructing young scholars in citation practices ranked in the accuracy of their citations while writing in an academic, professional publication. Genzinger and Wills analyzed 472 randomly selected citations from the three journals. Of those sampled, 122 (25.8%) had errors. The error rate per journal ranged from 21.7 to 28.7%. Of the 36 previous studies noted, the average error

rate range fell between 25 and 40%, placing Genzinger and Wills' study on the low end of the range.

It is surprising that a scholarly journal with such a prestigious reputation could publish so many errors, especially when a journal's reputation depends so heavily on the reader's ability to trust it as reliable (Rivkin, 2020). As part of their 2017 study, Genzinger and Wills (2017) contacted the editors of the journals to see what they had to say about the high citation error rates in their publications. Journal 1 placed the responsibility of citation accuracy solely on the author, though they admitted the editor did check all the citations and that editorial board members check citations when they review articles. Journal 2 also placed the responsibility of the citations on the author, explaining the editors noted incomplete citations and asked for corrections. Journal 3 explained that because their authors were library professionals, the editors trusted they had diligent citation practices. There have been no other reviews beyond those mentioned here. Genzinger and Wills noted that because they found no statistical significance between the error rates of the three journals, it appeared as though the error rates were unaffected by the varied methods.

One of the more recent developments in the academic writing arena is the use of citation management software. Rivkin (2020) characterized this as one of the first tools institutions or librarians provide for budding scholars. The authors added that reference management software programs such as EndNote, Mendeley, Zotero, or EasyBib are likely to decrease citation errors but cannot completely eliminate them. Sometimes these applications cannot collect all of the necessary information and require manual entry from the author. Much like a librarian reviewing a student's reference list, Rivkin explained citation managers can only check for proper citation format, and even then, they require the user to specify the proper style and data to use. According to Rivkin, these applications rely on the user to follow style rules, such as knowing which URL

to list in a reference. This means choosing the URL most likely to provide continual access to a resource, not just the one by which the author found the article (Homenda, 2021).

Penders (2018) noted citation practices also vary based on context, such as discipline, department, or even laboratory. Jomaa and Bidin (2019) explained this variation makes it impossible to create general principles that can be universally applied, adding citation principles even change depending on the style guide an author has adopted. For example, APA (2020) and Chicago styles (University of Chicago Press Staff, 2017) mandate transcribing the URL exactly the way it appears on the webpage. MLA (2021) allows the author to omit the protocol (http or https) and advises authors to shorten any URL that spans more than three lines. MLA specifically discourages the use of URL shorteners like bit.ly, but the guide instructs authors to include the homepage and possibly one further directory beyond that of the resources found (e.g., example.com/example). Conversely, Turabian et al. (2018) allow the author to simply cite the database where they found the resource but then suggest choosing the URL that will be most reliably accessed.

Style guides can even introduce new standards with each new edition. MLA did not advise about shortening links longer than three full lines in the eighth edition of their style guide (MLA, 2016). They added this instruction in the ninth edition (MLA, 2021). The sixth edition of the APA style guide advised authors to add the ftp:// protocol to URLs (APA, 2010). This technology has since been deemed insecure and has fallen out of popularity (“Chrome 86,” 2020). As a result, the seventh edition of the APA style guide omits this convention (APA, 2020). Similarly, the 15th edition of the Chicago style manual includes mention of ftp references (University of Chicago Press Staff, 2003), but the 17th edition excludes these protocols (University of Chicago Press Staff, 2017). Another depreciated web artifact involves the http protocol that receives much attention while no mention is made of the https protocol used by

over 80% of internet users (Electronic Frontier Foundation, 2022) and required by law for government websites ("The HTTPS-Only Standard," 2015). These antiquated terms were subsequently removed in the 17th edition of the Chicago style manual (University of Chicago Press Staff, 2017).

All these changes make it unsurprising that citation errors occur so often, but it remains unclear who is responsible for rectifying the issue. Rivkin (2020) explained that the culture of academia places the burden of training young scholars on librarians. The author added that academics believe academic journals bear some of the responsibility for curating the professional repository in which they claim to participate. According to Rivkin, journals place the blame solely on the authors. The inventor of the World Wide Web, Tim Berners Lee, has always placed the burden of keeping links active on the webmaster of the site hosting them (Krol & Zdonek, 2020).

Few dissenters resist the notion that academic librarians should continue to be a source of information and training for novice academics (Homenda, 2021). Rivkin (2020) asserted librarians must emphasize the importance of knowing the appropriate style guide and relying less on imprecise applications such as reference managers. Rivkin continued that although authors should always review and verify their references, some have suggested journals require them to provide an affidavit at the time of submission that all citations have been checked against the original content and provide full-text copies of all cited articles. Jones noted another popular suggestion, which involves pairing novice authors with more veteran members of a department to provide them with practical training and support.

Retracted Publications

Citation represents an ethical practice (Rubbo et al., 2019), so all researchers should concern themselves with proper citation (Penders, 2018), and consequences must exist to combat

any unethical behavior. It is the duty of the academic community to report publications that present falsified data or misrepresent their sources (Rubbo et al., 2019). Rubbo et al. (2019) suggested the consequence of having the publication retracted. A retraction significantly impacts the researcher and those associated with the publication because the publisher issues a formal notice, and those involved with the research process become associated with a clear waste of resources. The authors pointed out that some view retraction as an article's end of life, but Rivkin (2020) contended several studies have shown that retracted articles remain actively cited, adding most of the uses include a positive perspective on the retracted research.

Rubbo et al. (2019) conducted a review that revealed that retraction had little impact on an article's future performance. Rubbo et al. explained this use of the article continues because many researchers remain unaware of the retraction. All authors should ensure that the references they cite have not been amended or retracted, but this can be difficult (Rivkin, 2020). Rubbo et al. explained databases holding scholarly publications are not always updated to reflect an article's retraction status, thus perpetuating its use. Rubbo et al. added that the continued use of retracted articles compromises the works that are founded on that research. The authors contended this type of sloppy scholarship taints the studies relying on the retracted research as well as the institutions that financially support research and the governments that use these studies to draft public policies.

Peer Review

Hassan and Serenko (2019) contended that scholarly research does not end with a manuscript. Instead, they asserted, the process continues through the procedure of empirical verification. Peer review is considered the gold standard for research evaluations (Aksnes et al., 2019). Rivkin (2020) observed that the community relies on peer-reviewed articles in academic journals to stay current on knowledge and recent changes in their area of discipline. The author

explained that peer reviewers and editors should conduct a random reference verification during the manuscript review process to ensure accuracy and reliability.

Scholarship Leadership

Researchers have also examined the role institutions play in the equation. Research contaminated by link rot and improper citations wastes valuable resources (Rubbo et al., 2019), and it can also affect the reputation of the institution associated with the researcher (Zhao & Qiao, 2017). One elementary issue involves the number of dead links on an institution's website (Mills et al., 2021). This detracts from the institution's credibility (Krol & Zdonek, 2020). Additionally, many universities receive funding correlated to the reputation of their employees (West et al., 2017). Research output also represents a key component in college and university rankings (Zhao & Qiao, 2017). The traditional perspective is that research funding leads directly to innovation and knowledge (Rozell, 2020), but when science becomes a financial vehicle, vulnerabilities tend to emerge and can result in weak scholarship (Rivkin, 2020).

The use of research output for employment and promotion represents more than a matter of etiquette (West et al., 2017). Research functions as an important component in the reviewing of university faculty members (Zhao & Qiao, 2017). Many have questioned whether tenure or lifetime appointments should remain so intertwined with faculty's research output (Jones, 2021). The connection between job security and publication has pushed researchers away from high-risk research and toward more basic and short-term projects (Rozell, 2020). This way of valuing research has ethical and practical consequences that affect scholarship (Rivkin, 2020).

The most popular remedy to the lack of education for young scholars involves assuring that postgraduate students have access to sufficient guidelines and receive enough training to enable them to properly contribute scholarly work to their academic community (Jomaa & Bidin, 2019). Homenda (2021) asserted this responsibility still falls on the library and added young

scholars must receive training on persistent links. If an institution does not use some form of persistent URLs, the author contended, its leaders should begin investigating the process of adopting them immediately. Offering citation management software still represents good practice, but researchers must understand the various elements that make up a citation and how to manually correct them (Milewski et al., 2017).

Krol and Zdonek (2020) asserted the final recommendation for institutions is regular website maintenance. This includes a link audit on a regular basis. During any technological change, the authors explained, institutions must maintain the availability of resources. Robinson et al. (2018) asserted that the objective is to make it straightforward for researchers to distribute and utilize data, prioritizing the preservation of and access to data.

Citations Index

The concept of the citation index evolved to help researchers "understand the relationship between citing and cited works" (Hassan & Serenko, 2019). Aksnes et al. (2019) explained that a citation index lists the articles registered to the index. The authors continued that the index then counts how frequently that work is cited in future works submitted to the database. A citation index allows researchers to look up reference lists, also called reference chaining, to learn about the state of the topic they are researching (Colavizza et al., 2018).

Aksnes et al. (2019) described the field of bibliometrics as firmly linked to the birth of citation indexes. The authors explained that bibliometrics was created to study information retrieval, helping researchers identify pertinent articles in the vast archive of research literature. Researchers often confess to being swayed by a paper's citation index, leaning toward more prominent articles and overlooking those with fewer citations (Rivkin, 2020).

Since Eugene Garfield created the Science Citation Index, various versions of the citation index have emerged. Perhaps the most notable is the Web of Science, which includes many

different fields (Aksnes et al., 2019). Other notable citation indexes include Google Scholar, Science Direct, and Scopus from Elsevier (Bradshaw et al., 2021). Though the creation of these indexes has improved the coverage of literature, a large gap still persists in the humanities and in several social sciences (Aksnes et al., 2019). The fact these indexes exclude certain texts such as editorials, letters, and book reviews provides one possible explanation for this gap (West et al., 2017). While the Web of Sciences and Scopus generally include more articles from the sciences, non-subject-specific indexes such as Google Scholar yield a greater number of results for the humanities (Aksnes et al., 2019). Colavizza (2018) suggested that it might be necessary to create domain-specific indexes to help enrich those other catalogs with citation data.

In addition to the scope of the indexes themselves, the longevity and continuity of access to them represent one of the largest concerns about them. The academic publishing conglomerate Elsevier owns and operates Scopus and Science Direct (Mills et al., 2021). Over the last decade, Elsevier has come under fire for their contract negotiating tactics, which has resulted in a large number of research universities severing ties with them (MIT Libraries, 2018). This break has meant a loss of access to numerous citation indexes.

Similarly, the trajectory of the parent company Google has jeopardized access to publicly available citation index Google Scholar. Krol and Zdonek (2020) reported that Google has a reputation for shutting down projects once they no longer find them useful to their mission. For example, the authors described an incident where Google marketed Google+ to educators as a way to collaborate and share information but then shut down the service in April 2019, causing 43,000 Wikipedia links to go dead overnight. Google has shut down other services used by academics (e.g., Google Reader and Google Bookmarks). Disabling a service such as Google Scholar would equate to disrupting a spiderweb, destroying pathways and bridges to a multitude of scholarly connections (Król & Zdonek, 2020).

Persistent Identifiers

Academics have come to understand that the dynamic nature of the internet makes the availability of reliable links to web resources no trivial matter (Klein & Balakireva, 2021). Technologists, librarians, and scholars have continually developed methods and systems to preserve the sum of human knowledge (Robinson et al., 2018). Klein and Balakireva (2021) asserted persistent identifiers, such as DOIs, have been promoted as a reliable means of addressing the issues of link rot and proper citation indexing. The authors explained that the principle behind the DOI is that it will remain unchanged, providing access to the original resource even as its location changes. In that way, Klein and Balakireva claimed, DOIs also act as a citation index of sorts, mapping the use of resources. DOIs have become an essential part of the scholarly research process.

While DOIs have established themselves as the go-to standard for persistent identifiers of scholarly resources, other persistent identifiers have been introduced over the years (Klein & Balakireva, 2021). Some popular persistent link services include Handle, ARK, PURL, Perma.cc, and ArXiv (Homenda, 2021). Perma.cc, which emerged from research conducted at Harvard Law School's Berkman Klein Center for Internet and Society, provides a resource's full archive at the point when it is added to their service (Callister, 2021). This contrasts the index methods used with citation indexes and DOIs because it involves a hosted version of each work.

The persistent identifier arena continues to grow, but legal issues may arise for link service companies that archive a resource that does not belong to them (Callister, 2021). Although Elsevier's citation indexes hold the copyrights to the content they index (Mills et al., 2021), Callister (2021) explained service companies such as Perma.cc expect to stand on the established fair use and library exemptions. Callister added that the potential for risk exists

because Perma.cc's strategy has yet to be tested in court. Hence, those choosing a preservation method must consider the longevity of archival web services.

Homenda (2021) declared that proper citation practice means choosing the URL most likely to provide continuing access to a resource. One issue with DOIs involves the fact they rely on the expectation that a resource's publisher will keep their databases updated with a resource's location after it has changed (Klein & Balakireva, 2021). Homenda (2021) explained this can be difficult for smaller and institutional publishers because of the level of commitment it takes to maintain the URLs over time. Homenda pointed out this effort requires policies, technology, labor, and a large financial investment on the part of the publisher.

In another difficulty presented by DOIs, Klein and Balakireva (2021) noted that studies have continually shown authors do not utilize them when they should, despite their being a widely adopted and accepted best practice. The researchers added that many authors also choose to use the URL of the resource's landing page, even when a DOI exists. This compounds the problem of link persistence because it duplicates the work of finding the article once on the publisher's site. Homenda (2021) pointed to a recent survey where librarians from Indiana University Libraries examined what persistent link methods members of the Digital Library Federation's 195 members used as of August 2019. They found that only 49% of the artifacts in their digital collection claimed to have persistent links. Of those 49%, 37% had no discernable way to determine which technology they were using or if the links could resolve.

Klein and Balakireva (2021) described another troubling trend revealed in studies showing that publishers do not reliably resolve requests against DOIs. The authors added the resolution method might vary based on the user's choice of browser. The resolution method has also been shown to change over time.

According to Klein and Balakireva (2021), institutions have a comparable concern with consistency. In a recent study, 51.7% of DOI requests from outside of the institution's internal network failed. The authors explained this likely resulted from a certain licensing agreement with certain publishers. Just as surprising, however, the authors noted that just over 33% of DOIs from inside the institution's network also failed. This raises some legitimate concerns regarding the longevity of these persistent identifiers.

A contributing factor to the misuse of persistent identifiers involves the fact proper procedure can change from one style guide to another. The *Associated Press Stylebook* makes no mentioned of persistent identifiers at all (Associated Press, 2020). Turabian style emphasizes using the most reliable link for a resource and also specifically mentions using DOIs, noting that the author should append all DOIs to "https://doi.org/" (Turabian et al., 2018). Chicago style echoes that DOIs are more reliable than traditional URLs and that authors should precede the DOI with "https://doi.org/:" (University of Chicago Press Staff, 2017). MLA (2021) repeats the same advice. APA (2020) follows the same guidelines but adds the use of http://dx.doi.org or "doi:" as an alternative format. The International DOI Foundation (2016) found this questionable as it conflicts with DOI's governing body's guidelines for formatting their resources. The guidelines asked that the https://doi.org/ formula be used exclusively.

Something created to be a persistent reference should be consistent over time, but as with citations, style guides' suggested DOI formatting has changed over time. The eighth edition of the MLA handbook instructed the author to cite a DOI simply by adding "doi:" in front of the article's DOI (MLA, 2016). The sixth edition of the APA (2010) style guide and the 15th edition of *The Chicago Manual of Style* (Staff, 2003) suggested the same formatting.

All the current editions of the mentioned style guides suggest replacing a URL with a DOI if possible. Notably, not all scholarly works will have a DOI to use (Klein & Balakireva,

2021). This follows the philosophy that the link most likely to be continuously reliable should be used in a citation (Homenda, 2021). For a few years, some have argued for making it common practice to provide both the original URL and a DOI (Homenda, 2021). Researchers should conduct further investigations to examine this method's efficacy.

Research Gaps

Kaufman and Campana (2019), Krol and Zdonek (2020), Bradshaw et al. (2021), and Klein and Balakireva (2021) have all conducted extensive research into the nature and cause of link rot. Homenda (2021) and Klein and Balakireva (2021) focused closely on the use of persistent modifiers, and Callister (2021) warned of the legal implications of certain services. Genzinger and Wills (2017), Rubbo et al. (2019), and Rivkin (2020) all broadly examined the way citation accuracy impacts scholarly research, but a few gaps remain in the whole of the research.

One area in need of further examination involves the need for research comparing rates across disciplines. Aksnes et al. (2019) delved into the lifecycle and norms of citations over a few fields. In their broader study of link rot, Krol and Zdonek (2020) analyzed the problem at large but did not break the issue down by area of study. A significant portion of the research that did address the subject matter focused specifically on one discipline. Law (Zittrain et al., 2013), health sciences (Rivkin, 2020), and information science (Homenda, 2021) represented some of the most common areas covered.

Even fewer studies breaking down the use of DOIs by disciplines existed. Homenda (2021) analyzed the use of persistent links by institution. Klein and Balakireva (2021) revealed the number of DOIs that did not link properly and how the rates differed from inside and outside the institutional network. To date, no researchers have looked at link rot or the reliability of DOIs across disciplines.

III. METHODOLOGY

The purpose of this descriptive study was to measure, report, and analyze the effect of link rot on scholarly research with the intention of informing university policy on scholarly publishing and the integrity of institutional research. The essential elements associated with the study's methodology appear in the following sections.

Research Design and Methodology

A nonexperimental, quantitative research design was used to address the study's topic (Edmonds & Kennedy, 2016). The specific research methodology was retrospective causal comparative (Fraenkel et al., 2018). The main benefit of causal-comparative research, or the ex post facto method, is that the design allows for an evaluation of variables that should not be manipulated as the event under study has already occurred. Importantly, the causal-comparative research method also allows the researcher to determine how a particular phenomenon occurred (Adams & Lawrence, 2018). Study data were achieved archivally from the professional literature and were delimited to five major academic domains of professional publications over a 10-year period.

Study Procedures

Historical archived data were compiled and analyzed using a self-created tool to evaluate the extent to which a publication had been affected by link rot. A total of 2,500 peer-reviewed scholarly articles were collected for the purpose of this study. Articles with publication dates between 2010 and 2020 were selected from a variety of journals and other scholarly publishing

outlets. The publications selected represented one of five subject matters: (a) arts and humanities; (b) business; (c) health and medicine; (d) science, math, and technologies; and (e) social sciences. Each subject matter was represented by 500 articles, and each year in the 10-year range was represented by 250 publications.

A command-line script was run on all publications. The script mined the publications for references to online resources, and these references were tested for connection to their original source. The response to the connection request was recorded. The request either succeeded or returned an error code, and all codes were recorded and organized by code. Due to the nature of this study, the tool also specifically indicated any links directed to a DOI or arXiv source. The source code appears at https://github.com/rottingresearch/linkrot.

Parameters were established in order to avoid any bias or inconsistent judgment and to standardize the results. Email addresses were noted but not checked for active status due to the spam risk. Redirects were treated as broken links. Resources requiring credentials to view an abstract or details were regarded as broken links. No navigation past the initial URL entry was allowed for verification. Every error was validated and checked for accuracy with a manual examination of every link returned with an error code.

All publications and resources were gathered from archived and historical data. Although confidentiality did not present a concern, measures were taken to remove any possibility of bias. All articles were categorized into their specified subject area and labeled with a random number between 1 and 500 and the year they were published. The output of the command-line script contained no identifying data. This output only included a list of all links mined from the publication, the type of resources, successful link requests, and failed requests with the error code returned.

Metrics that were analyzed included percentage of publications affected, average number of links broken, and breakdown of errors by type. Furthermore, these results were used to sort by publication year and subject matter.

Research Questions and Hypotheses

Four formally stated research questions and hypotheses addressed the study's topic and research problem.

1. To what degree is link rot evident in scholarly research?

   H_0: There will be no statistically significant degree of link rot evident in scholarly research.

2. To what degree is link rot evident for DOIs?

   H_a: There will be no statistically significant degree of link rot for DOIs evident in scholarly research.

3. Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for link rot?

   H_a: There will be a statistically significant effect for academic domain on the percentage of broken links.

4. Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for DOI link rot?

   H_a: There will be a statistically significant effect for academic domain on the percentage of broken DOI links (DOI link rot).

Data Analysis

Statistical Power Analysis

Statistical power analysis was conducted using the G*Power statistical software Version 3.1.9.2 for sample size estimation purposes to detect statistically significant findings in the

study's analyses. An alpha level of $p = .05$ and power index (1–β) of .80 served as the parameters of power calculation for respective statistical techniques. In Research Questions 1 and 2, an anticipated medium effect ($d = .50$) required a sample size of 27 to detect a statistically significant finding and a sample size of 12 for an anticipated large effect ($d = .80$) for the use of the one-sample *t* test in both research questions. In Research Questions 3 and 4, an anticipated medium effect ($f = .25$) required a sample size of 200 to detect a statistically significant finding and a sample size of 80 for an anticipated large effect ($f = .40$) in the 1 x 5 ANOVA used in both research questions (Faul et al., 2009).

Foundational Descriptive Statistical Analyses

Study data were analyzed at the preliminary foundational level using descriptive statistical techniques. The study's demographic identifying information was addressed using the descriptive statistical techniques of frequencies and percentages. Study data associated with the phenomenon of broken links were addressed using the descriptive statistical techniques. Specifically, frequencies, measures of central tendency (i.e., mean scores), variability (i.e., minimums, maximums/standard deviations), standard errors of the mean, and measures of data normality (i.e., skewness, kurtosis) represented the descriptive statistical techniques used in the analyses.

Analyses by Research Question and Hypothesis

The statistical significance of mean score findings in Research Questions 1 and 2 was addressed using the one-sample *t* test. The assumption of data normality associated with the use of the one-sample *t* test was addressed by evaluating respective skew and kurtosis values. The conventions of data normality espoused by George and Mallery (2019) were specifically applied to the dependent variables in Research Questions 1 and 2 for assumption testing purposes. The probability level of $p \leq .05$ served as the threshold value for statistically significant one-sample *t*

test findings in Research Questions 1 and 2. Cohen's *d* was used as the measure of effect size for both Research Questions 1 and 2. The conventions of effect size interpretation proposed by Sawilowsky (2009) were used for these questions as well.

In Research Questions 3 and 4, a one-way analysis of variance (1 x 5 ANOVA) was used to assess the statistical significance of overall effect exerted by the respective independent variable upon the dependent variable in each research question. The assumption of data normality associated with the use of ANOVA was addressed by evaluating respective skew and kurtosis values. The conventions of data normality espoused by George and Mallery (2020) were specifically applied to the dependent variables in Research Questions 3 and 4 for assumption testing purposes. The assumption of homogeneity of variances was evaluated by interpreting the respective Levene *F* values in each research question. Follow-up post hoc analyses were conducted specifically using Tukey's honest significant difference (HSD) test. The probability level of $p \leq .05$ was selected as the threshold value for statistically significant ANOVA findings in Research Questions 3 and 4. The n^2 test statistic was used as the measure of effect size for both research questions, as were the conventions of effect size interpretation proposed by Sawilowsky (2009).

Study data were analyzed and reported using the 28th version of IBM's SPSS. Chapter 4 contains the formal reporting of findings achieved in the study.

IV. RESULTS

The purpose of this descriptive study was to measure, report, and analyze the effect of link rot on scholarly research with the intention of informing university policy on scholarly publishing and the integrity of institutional research. The study was delimited to five specific academic domains of inquiry. Study data were accessed through a content analysis process of the professional literature. A causal-comparative research design was used to address the study's topic and research problem. Four formally stated research questions and hypotheses addressed the study's purpose. Descriptive and inferential statistical techniques were used to analyze study data at the preliminary, foundational level and for the four research questions and hypotheses.

Preliminary Descriptive Statistical Findings

The study's preliminary findings were achieved through the application of descriptive statistical analyses. The following sections report the findings achieved at the study's preliminary foundational level.

Demographic Information

The study's primary demographic information was evaluated using descriptive statistical techniques of frequencies and percentages. Table 1 presents a summary of finding for the descriptive statistical analysis of the study's demographic identifying information for data pertaining to academic domain and year of data access.

Table 1

Descriptive Statistical Summary: Demographic Variables (Academic Domain; Year)

Variable	*n*	%	Cumulative %
Academic domain			
Arts and humanities	500	20.00	20.00
Business	500	20.00	40.00
Health and medicine	500	20.00	60.00
Science/math/technologies	500	20.00	80.00
Social sciences	500	20.00	100.00
Missing	0	0.00	100.00
Year			
2013	250	10.00	10.00
2014	250	10.00	20.00
2015	250	10.00	30.00
2016	250	10.00	40.00
2017	250	10.00	50.00
2018	250	10.00	60.00
2019	250	10.00	70.00
2020	250	10.00	80.00
2021	250	10.00	90.00
2022	250	10.00	100.00
Missing	0	0.00	100.00

Initial Descriptive Statistical Findings: Errors by Academic Domain

Descriptive statistical techniques of frequencies, measures of typicality, variability, standard errors of the mean, and data normality were utilized to assess the study's response set data for errors by academic domain. Table 2 presents a summary of finding for the descriptive statistical analysis of the study's response set data for errors by academic domain. Inaccessible (i.e., broken) links are characterized by the HTTP status code they return. A return code in the 300s indicates a redirection; an error code in the 400s indicates an error that appears to have been caused by the client, and an error code in the 500s indicates the server failed to fulfill a request.

Table 2

Descriptive Statistics Summary: Errors by Academic Domain

Academic domain Errors	*M*	*SD*	*n*	SE_M	Min	Max	Skew	Kurtosis
Arts and humanities								
300	0.06	0.67	500	0.03	0.00	10.00	13.66	195.61
400	1.24	2.95	500	0.13	0.00	30.00	5.38	39.58
500	1.18	4.77	500	0.21	0.00	90.00	13.77	241.85
Business								
300	0.06	0.51	500	0.02	0.00	10.00	15.94	292.18
400	1.61	2.66	500	0.12	0.00	27.00	3.49	19.95
500	3.52	8.32	500	0.37	0.00	85.00	6.08	44.44
Health and medicine								
300	0.24	1.03	500	0.05	0.00	10.00	5.72	37.38
400	1.62	4.18	500	0.19	0.00	47.00	5.34	39.23
500	0.59	1.64	500	0.07	0.00	19.00	5.57	44.69
Science/math/technologies								
300	0.03	0.19	500	0.009	0.00	2.00	7.09	54.92
400	1.46	3.36	500	0.15	0.00	31.00	4.36	23.95
500	4.11	13.13	500	0.59	0.00	90.00	4.22	19.25
Social sciences								
300	0.09	0.46	500	0.02	0.00	7.00	8.71	105.24
400	5.24	7.52	500	0.34	0.00	41.00	1.97	3.84
500	3.85	6.97	500	0.31	0.00	46.00	2.45	7.39

Initial Descriptive Statistical Findings: Broken Links by Academic Domain

Table 3 presents a summary of finding for the descriptive statistical analysis of the study's response set data for broken links by academic domain.

Table 3

Descriptive Statistics Summary: Broken Links by Academic Domain

Academic Domain	*M*	*SD*	*n*	SE_M	Min	Max	Skew	Kurtosis
Arts and humanities								
Broken link %	0.39	0.33	358	0.02	0.00	1.00	0.46	-0.81
Business								
Broken link %	0.39	0.26	472	0.01	0.00	1.00	0.13	-0.52
Health and medicine								
Broken link %	0.32	0.30	423	0.01	0.00	1.00	0.61	-0.50
Science/math/technologies								
Broken link %	0.34	0.30	350	0.02	0.00	1.00	0.51	-0.81
Social sciences								
Broken link %	0.37	0.25	426	0.01	0.00	1.00	0.20	-0.40

Initial Descriptive Statistical Findings: Broken DOI Links by Academic Domain

Table 4 presents a summary of finding for the descriptive statistical analysis of the study's response set data for broken DOI links by academic domain.

Table 4

Descriptive Statistics Summary: Broken DOI Links by Academic Domain

Variable	*M*	*SD*	*n*	SE_M	Min	Max	Skewness	Kurtosis
Arts and humanities								
Broken DOI %	0.42	0.43	160	0.03	0.00	1.00	0.34	-1.59
Business								
Broken DOI %	0.44	0.44	353	0.02	0.00	1.00	0.25	-1.70
Health and medicine								
Broken DOI %	0.33	0.39	222	0.03	0.00	1.00	0.73	-0.98
Science/math/technologies								
Broken DOI %	0.23	0.29	220	0.02	0.00	1.00	1.11	0.28
Social sciences								
Broken DOI %	0.40	0.32	299	0.02	0.00	1.00	0.12	-1.24

Findings by Research Question

The study's four research questions and hypotheses were addressed using descriptive and inferential statistical techniques. The probability level of $p \leq .05$ served as the threshold value for statistical significance of findings achieved in the study's analyses. Effect sizes achieved in the analyses were interpreted using the conventions of effect size interpretations offered by Sawilowsky (2009). The analysis of study data and reporting of findings was conducted using IBM's 28th version of SPSS. The following sections present the findings achieved according to the study's five stated research questions.

Research Question 1

The first research question was: To what degree is link rot evident in scholarly research? The one-sample *t* test was used to assess the statistical significance of mean difference between the overall broken link value and the test value of .50. The assumption of data normality was addressed statistically using the data array's skew and kurtosis values. The data array's skew value of 0.39 and kurtosis value of -0.61 fell well within the parameters of data normality proposed by George and Mallery (2019) for skew (-/+2.0) and kurtosis (-/+7.0). Consequently, these values satisfied the assumption of normality.

The percentage of overall broken links identified in the data set was 36%. The extent of overall broken links was statistically significantly less than the test value of 50% ($t_{(2028)} = -21.53$, $p < .001$). The magnitude of effect for the difference in broken link percentage in the study's data set and the test value of .50 was considered medium ($d = .48$). Table 5 presents a summary of finding for the statistical significance of the study's overall broken link percentage in comparison to the stated test value of .50.

Table 5

One-Sample t *Test Summary: Overall Broken Link Comparison*

Variable	*M*	*SD*	μ	*t*	*p*	*d*
Overall broken links	0.36	0.29	0.50	-21.53	< .001	0.48

Note. Degrees of freedom for the *t* statistic = 2028. *d* represents Cohen's *d.*

Hypothesis

The hypothesis for Research Question 1 was: There will be no statistically significant degree of link rot evident in scholarly research. Considering the statistically significant finding for the difference in the study's data set's overall percentage of broken links, the null hypothesis in Research Question 1 was rejected.

Research Question 2

The second research question was: To what degree is link rot evident for DOIs? The one-sample *t* test was used to assess the statistical significance of mean difference between the overall broken link value for DOIs and the test value of .50. The assumption of data normality was addressed statistically using the data array's skew and kurtosis values. The data array's skew value of 0.51 and kurtosis value of -1.25 fell well within the parameters of data normality proposed by George and Mallery (2019) for skew (-/+2.0) and kurtosis (-/+7.0). As a result, these values satisfied the assumption of normality.

The percentage of DOI broken links identified in the study's data set was 37%. The extent of DOI broken links was statistically significantly less than the test value of 50% ($t_{(1253)}$ = -11.94, $p < .001$). The magnitude of effect for the difference in broken link percentage in the study's data set and the test value of .50 was considered between small and medium ($d = .34$). Table 6 presents a summary of finding for the statistical significance of the study's DOI broken link percentage in comparison to the stated test value of .50.

Table 6

One-Sample t *Test Summary: DOI Broken Link Comparison*

Variable	*M*	*SD*	μ	*t*	*p*	*d*
Broken DOI links	0.37	0.39	0.5	-11.94	< .001	0.34

Note. Degrees of freedom for the *t* statistic = 1253. *d* represents Cohen's *d.*

Hypothesis

The hypothesis for Research Question 2 was: There will be no statistically significant degree of link rot for DOIs evident in scholarly research. Considering the statistically significant finding for the difference in the data set's percentage of DOI broken links, the null hypothesis in Research Question 2 was rejected.

Research Question 3

The third research question was: Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for link rot? A 1 x 5 ANOVA was used to assess the overall statistical significance of effect for academic domain and the percentage of link rot. The assumptions of data normality and absence of influential outliers were satisfied. However, the assumption of homogeneity of variances was violated (F[2, 2024] = 13.85, $p < .001$). The one-way ANOVA is robust against violations of homogeneity of variables when data are relatively normally distributed and influential outliers are not present in the analysis (Field, 2018). As a result, the analysis proceeded using the one-way ANOVA and post hoc testing.

The ANOVA analysis produced a statistically significant finding (F[4, 2,024] = 5.17, $p <$ .001), indicating there were significant differences in percent of broken links among the levels of academic domain (see Table 7). The eta squared was 0.01, indicating academic domain

explained approximately 1% of the variance in the percentage of broken links. The means and standard deviations of the ANOVA analysis appear in Table 8.

Table 7

Analysis of Variance Summary: Effect of Academic Domain Upon Percentage of Broken Links

Model	*SS*	*df*	*F*	*p*	η_p^2
Academic domain	1.71	4	5.17	< .001	0.01
Residuals	167.16	2024			

Table 8

Mean, Standard Deviation, and Sample Size for Percentage of Broken Links by Academic Domain

Academic Domain	*M*	*SD*	*n*
Arts and humanities	0.39	0.33	358
Business	0.39	0.26	472
Health and medicine	0.32	0.30	423
Science/math/technologies	0.34	0.30	350
Social sciences	0.37	0.25	426

Follow-Up Post Hoc Analyses

Follow-up post hoc analyses using paired *t* tests were calculated between each pair of measurements to further evaluate the differences among the variables. Tukey HSD *p*-value adjustments were specifically used to correct for the effect of multiple comparisons on the family-wise error rate. For the main effect of academic domain, the mean percentage of broken links for arts and humanities ($M = 0.39$, $SD = 0.33$) was statistically significantly greater than for the domain of health and medicine ($M = 0.32$, $SD = 0.30$, $p = .009$). For the main effect of academic domain, the mean percentage of broken links for business ($M = 0.39$, $SD = 0.26$) was statistically significantly greater than for the domain of health and medicine ($M = 0.32$, $SD = 0.30$, $p = .001$). For the main effect of academic domain, the mean percentage of broken links for business ($M = 0.39$, $SD = 0.26$) was statistically significantly greater than for the domain of science/math/technologies ($M = 0.34$, $SD = 0.30$, $p = .04$).

Hypothesis

The hypothesis for Research Question 3 was: There will be a statistically significant effect for academic domain upon the percentage of broken links. Considering the statistically significant overall effect of academic domain on percentage of broken links, the alternative hypothesis in Research Question 3 was retained.

Research Question 4

The fourth research question was: Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for DOI link rot? A 1 x 5 ANOVA was used to assess the overall statistical significance of effect for academic domain and the percentage of DOI link rot. The assumptions of data normality and absence of influential outliers were satisfied. However, the assumption of homogeneity of variances was violated ($F[4, 1249] = 46.90$, $p < .001$). The one-way ANOVA is robust against violations of

homogeneity of variables when data are relatively normally distributed and influential outliers are not present in the analysis (Field, 2018). As a result, the analysis proceeded using the one-way ANOVA and post hoc testing.

The resultant ANOVA analysis finding was statistically significant (F[4, 1,249] = 12.72, $p < .001$), indicating significant differences existed in percent of broken DOI links among the levels of academic domain (see Table 9). The eta squared was 0.04, indicating academic domain explained approximately 4% of the variance in the percentage of broken DOI links. The means and standard deviations of the ANOVA analysis appear in Table 10.

Table 9

Analysis of Variance Summary: Effect of Academic Domain Upon Percentage of Broken DOI Links

Model	*SS*	*df*	*F*	*p*	η_p^2
Academic domain	7.30	4	12.72	< .001	0.04
Residuals	179.14	1249			

Table 10

Mean, Standard Deviation, and Sample Size for Percentage of Broken DOI Links by Academic Domain

Academic Domain	*M*	*SD*	*n*
Arts and humanities	0.42	0.43	160
Business	0.44	0.44	353
Health and medicine	0.33	0.39	222
Science/math/technologies	0.23	0.29	220
Social sciences	0.40	0.32	299

Follow-Up Post Hoc Analyses

Follow-up post hoc analyses were conducted using paired *t* tests. Tukey's HSD *p*-value adjustments were used to correct for the effect of multiple comparisons on the family-wise error rate. For the main effect of academic domain, the mean percentage of broken DOI links for arts and humanities ($M = 0.42$, $SD = 0.43$) was statistically significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p < .001$). For the main effect of academic domain, the mean percentage of broken DOI links for business ($M = 0.44$, $SD = 0.44$) was significantly greater than for the domain of health and medicine ($M = 0.33$, $SD = 0.39$, $p = .007$). For the main effect of academic domain, the mean percentage of broken DOI links for business ($M = 0.44$, $SD = 0.44$) was significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p < .001$).

For the main effect of academic domain, the mean percentage of broken DOI links for health and medicine ($M = 0.33$, $SD = 0.39$) was significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p = .03$). For the main effect of academic domain, the mean percentage of broken DOI links for science/math/technologies ($M = 0.23$, *SD*

= 0.29) was significantly lesser than for the domain of social sciences (M = 0.40, SD = 0.32, p < .001).

Hypothesis

The hypothesis for Research Question 4 was: There will be a statistically significant effect for academic domain upon the percentage of broken DOI links. Considering the statistically significant overall effect for academic domain upon percentage of broken DOI links, the alternative hypothesis in Research Question 4 was retained.

Summary

Study data were accessed through content analysis of professional literature spanning the 10 years between 2013 and 2022. Five specific academic domains in the scholarly literature were identified for study purposes. The overall percentage of link rot was statistically significantly lesser than the study's adopted threshold level of 50%. Similarly, the overall percentage of DOI link rot was statistically significantly lesser than the study's adopted threshold level of 50%.

Statistically significant overall effects for academic domain upon broken links and broken DOI links manifested in the analyses for Research Questions 3 and 4. The single greatest degree of link rot emerged in the academic domains of arts and humanities and business at 39%. The academic domain reflecting the least degree of link rot was health and medicine at 32%. The single greatest degree of DOI link rot emerged in the academic domain of business at 44%, closely followed by the academic domain of arts and sciences at 42%. The academic domain reflecting the least degree of DOI link rot was science/math/technologies at 23%.

Chapter 5 contains a thorough discussion of the findings reported in Chapter 4.

V. DISCUSSION

The purpose of this descriptive study was to measure, report, and analyze the effect of link rot on scholarly research with the intention of informing university policy on scholarly publishing and the integrity of institutional research. Link rot was generally defined as the phenomenon of resources ceasing to be accessible over time due to their relocation or their originally cited location becoming permanently unavailable. The HTTP status code returned by a link characterizes it as inaccessible or broken. A return code in the 300s indicated a redirection. An error code in the 400s indicated an error that appeared to have been caused by the client. An error code in the 500s indicated the server failed to fulfill a request. Command-line was generally defined as a computer interface that allowed tasks and processes to be executed via text strings. A script was generally defined as a sequence of computer code strung together to run a unified process or task. Mining was generally defined as searching through data to find a resource. A persistent identifier was generally defined as a long-lasting link to a resource that remained active despite any change in data structures. A publication was generally defined as a work that was made accessible to the public. *Robust* was generally defined as strong, rigorous, and steadfast.

Review of Methodology

Twenty-five hundred peer-reviewed scholarly articles were collected for the purpose of this study. Articles published between the years 2013 and 2022 were selected from a variety of journals and other scholarly publishing outlets in one of five subject areas: (a) arts and

humanities; (b) business; (c) health and medicine; (d) science, math, and technologies; and (e) social sciences. Each subject area was represented by 500 articles, and each year was represented by 250 publications.

A command-line script was run on all publications. The script mined the publications for references to online resources. These references were tested for connection to their original source, and the response from the connection request was recorded. The request either returned successfully or with an error code, and all codes were recorded and organized by code. Due to the nature of this study, the tool also specifically identified any links directed to a DOI or arXiv source. The source code for the tool appears at https://github.com/rottingresearch/linkrot.

Summary of Results

The researcher set parameters to avoid any bias or inconsistent judgment. As such, the following guidelines were established to standardize the results. Email addresses were noted but not checked for active status to avoid spam risk. Redirects were treated as broken links, and resources requiring credentials to view an abstract or details were regarded as broken links. No navigation past the initial URL entry was allowed for verification. Every error was validated and checked for accuracy by examining every link returned with an error code.

All publications and resources were gathered from archived and historical data. Although confidentiality presented no concern, measures were taken to remove any possibility of bias. All articles were categorized into their specified subject areas and labeled with their year of publication and the random assignment of a number between 1 and 500. The output of the command-line script contained no identifying data and included only a list of all links mined from the publication, the type of resource, successful link requests, and failed requests with the error code returned.

Analyzed metrics included percentage of publications affected, average number of links broken, and errors by type. These results were used to classify by publication year and subject matter. The percentage of overall broken links identified within the study was 36%. A broken link was defined as a link that returned a server response. This ranged from 44% in business publications to 23% in science, math, and technology. Arts and humanities had a broken link percentage of 42% compared to 40% for social science and 33% for health and medicine.

The percentage of DOI broken links identified in the data set was 37%. The extent of DOI broken links was statistically significantly less than the test value of 50% (t (1253) = -11.94, $p < .001$). The magnitude of effect for the difference in broken link percentage in the study's data set and the test value of 50% was considered between small and medium (d = .34)

For the main effect of academic domain, the mean percentage of broken links for arts and humanities (M = 0.39, SD = 0.33) was statistically significantly greater than for the domain of health and medicine (M = 0.32, SD = 0.30, p = .009). For the main effect of academic domain, the mean percentage of broken links for business (M = 0.39, SD = 0.26) was statistically significantly greater than for the domain of health and medicine (M = 0.32, SD = 0.30, p = .001). For the main effect of academic domain, the mean percentage of broken links for business (M = 0.39, SD = 0.26) was statistically significantly greater than for the domain of science/math/technologies (M = 0.34, SD = 0.30, p = .04).

For the main effect of academic domain, the mean percentage of broken DOI links for arts and humanities (M = 0.42, SD = 0.43) was statistically significantly greater than for the domain of science/math/technologies (M = 0.23, SD = 0.29, $p < .001$). For the main effect of academic domain, the mean percentage of broken DOI links for business (M = 0.44, SD = 0.44) was significantly greater than for the domain of health and medicine (M = 0.33, SD = 0.39, p = .007). For the main effect of academic domain, the mean percentage of broken DOI links for

business (M = 0.44, SD = 0.44) was significantly greater than for the domain of science/math/technologies (M = 0.23, SD = 0.29, $p < .001$). For the main effect of academic domain, the mean percentage of broken DOI links for health and medicine (M = 0.33, SD = 0.39) was significantly greater than for the domain of science/math/technologies (M = 0.23, SD = 0.29, $p = .03$). For the main effect of academic domain, the mean percentage of broken DOI links for science/math/technologies (M = 0.23, SD = 0.29) was significantly lesser than for the domain of social sciences (M = 0.40, SD = 0.32, $p < .001$).

Discussion by Research Question

Research Question 1

The first research question was: To what degree is link rot evident in scholarly research? The study showed the overall percentage of broken links was 36%, with a broken link defined as a link that returned a server response. This ranged from 44% in business publications to 23% in science, math, and technology publications. Arts and humanities publications had a broken link percentage of 42%, and 40% of social science publications and 33% of health and medicine publications returned a server response.

Given the low number of studies on broken links, researchers had yet to establish a threshold value for the percentage of broken links in the field of link rot. Unlike studies conducted by Genzinger and Wills (2017), Homenda (2021), and Rivkin (2020), which all focused on one discipline, this study took the broader view adopted by Krol and Zdonek (2020) and went one step further to include the breakdown across various areas of study.

Although this study fell short of its 50% threshold, the overall percentage of broken links (36%) and range of broken links by discipline (23–44%) aligns with previous studies such as those conducted by Krol and Zdonek (2020), Rivkin (2020), and Genzinger and Wills (2017). The percentage of broken links in this study fell well below the 70% rate estimated by Krol and

Zdonek (2020) for nonacademic publications. Given the alignment of broken link percentages calculated by discipline between this and previous studies, these results could help establish a concrete threshold for future research.

Research Question 2

The second research question was: To what degree is link rot evident for DOIs? The study's data showed 37% of DOI links were broken. The extent of broken DOI links was statistically significantly less than the test value of 50% ($t_{(1253)} = -11.94$, $p < .001$). The magnitude of effect for the difference in broken link percentage between the study's data set and the test value of 50% was considered between small and medium ($d = .34$)

The overall percentage of broken DOI links is especially interesting, given that it exceeded the overall percentage of broken links for all URLs. Developers created the DOI to address issues with resource availability (Klein & Balakireva, 2021), but it appears that DOIs are associated with a similar, if not slightly greater, percentage of inaccessible resources. Although notable, the high percentage of broken DOI links presents less of a surprise given that Klein and Balakireva (2021) showed 33% of DOI links to be broken on one institution's internal network. Klein and Balakireva found 51.7% of DOIs were unreachable from outside of the institution's network. This study did not address the effect location might have on access. Many issues contributed to the high percentage of broken DOI links in the Klein and Balakireva study. The authors explained that with DOIs, the publisher must take responsibility for updating the location up-to-date. Therefore, no inherent process ensures the consistent update of these links. The authors added some publishers may not have the resources or ability to make such updates, especially considering the heavy burden of doing so.

Some publishers have developed independent policies to address DOI longevity. Homenda (2021) asserted the best link to use is the one most likely to provide continual access to

the resource. However, some publishers have pointed their DOIs to pages behind a paywall without providing any means of checking the link besides logging in. This situation returns an HTTP response code in the 300s. The best practice in the case of paywalled resources is to point the link to the publicly viewable abstract with instructions on how to gain full access. This allows verification of the link to the resource and produces a successful HTTP response code.

Publishers have also compromised the integrity of DOI links by using a third-party service to handle their DOIs. Crossref serves as one example that appeared frequently in 300-level response codes. Crossref claims it "makes research objects easy to find, cite, link, assess, and reuse" (Hendricks, 2021). Using a third-party service adds another possible point of weakness in the process. A noticeable portion of the inaccessible links pointed to services like this.

Citation errors represented another visible cause of broken DOI links. As explained in Chapter 2, style guides have strict rules for how to cite DOI references, but these have changed over time. The study showed several links formatted using outdated guidelines, which made the links unresolvable without the manual interpretation of someone with knowledge of the formatting changes over time.

The results of this study should have a real impact on future research. Developers created DOIs to reduce link rot, yet their usage appears to have no effect on the percentage of broken links at best and perhaps a slightly negative effect at worst. Despite the popularity and usage of DOIs in research, education about them appears to be lagging. Future researchers might consider focusing on alternatives to DOIs or investigating different methods of storing and accessing digital information beyond permalinks. Some other prominent projects include ArXiv, Memento, and Perma.cc. This study initially included ArXiv links, but too few references emerged in the dataset to produce any meaningful data.

Research Question 3

The third research question was: Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for link rot? A 1 x 5 ANOVA was used to assess the overall statistical significance of effect for academic domain and the percentage of link rot. The results satisfied the assumptions of data normality and absence of influential outliers. However, they violated the assumption of homogeneity of variances ($F[2, 2024] = 13.85$, $p < .001$). The one-way ANOVA is robust against violations of homogeneity of variables when data are relatively normally distributed and influential outliers are not present in the analysis (Field, 2018). As a result, the analysis proceeded using the one-way ANOVA and post hoc testing.

The resultant finding for the ANOVA analysis was statistically significant ($F[4, 2{,}024] = 5.17$, $p < .001$), indicating significant differences existed in percent of broken links among the levels of academic domain (see Table 7). The eta squared was 0.01, indicating academic domain explained approximately 1% of the variance in the percentage of broken links. The means and standard deviations of the ANOVA analysis appear in Table 8.

Follow-up post hoc analyses using paired *t* tests were calculated between each pair of measurements to further evaluate the differences among the variables. Tukey HSD *p*-value adjustments were specifically used to correct for the effect of multiple comparisons on the family-wise error rate. For the main effect of academic domain, the mean percentage of broken links for arts and humanities ($M = 0.39$, $SD = 0.33$) was statistically significantly greater than for the domain of health and medicine ($M = 0.32$, $SD = 0.30$, $p = .009$). For the main effect of academic domain, the mean percentage of broken links for business ($M = 0.39$, $SD = 0.26$) was statistically significantly greater than for the domain of health and medicine ($M = 0.32$, $SD = 0.30$, $p = .001$). For the main effect of academic domain, the mean percentage of broken links for

business (M = 0.39, SD = 0.26) was statistically significantly greater than for the domain of science/math/technologies (M = 0.34, SD = 0.30, p = .04).

Significant differences in the percentage of broken links emerged across the disciplines addressed in this study. Future researchers should investigate this phenomenon to determine its causes. This study represents one of the first to compare the percentage of broken links across disciplines, so no past research exists against which to compare the results. Future researchers should reproduce this study to provide a more robust understanding of the percentage differences across disciplines and to determine why one discipline might have a higher or lower percentage of broken links than another.

Research Question 4

The fourth research question was: Considering the five academic areas of research identified for study purposes, is there an overall statistically significant effect for DOI link rot? A 1 x 5 ANOVA was used to assess the overall statistical significance of effect for academic domain and the percentage of DOI link rot. The results satisfied the assumptions of data normality and absence of influential outliers. However, they violated the assumption of homogeneity of variances (F[4, 1249] = 46.90, p < .001). The one-way ANOVA is robust against violations of homogeneity of variables when data are relatively normally distributed and influential outliers are not present in the analysis (Field, 2018). As a result, the analysis proceeded using the one-way ANOVA and post hoc testing.

The resultant ANOVA analysis finding was statistically significant (F[4, 1,249] = 12.72, p < .001), indicating significant differences existed in percent of broken DOI links among the levels of academic domain (see Table 9). The eta squared was 0.04, indicating academic domain explained approximately 4% of the variance in the percentage of broken DOI links. The means and standard deviations of the ANOVA analysis appear in Table 10.

Follow-up post hoc analyses were conducted using paired *t* tests. Tukey's HSD *p*-value adjustments were used to correct for the effect of multiple comparisons on the family-wise error rate. For the main effect of academic domain, the mean percentage of broken DOI links for arts and humanities ($M = 0.42$, $SD = 0.43$) was statistically significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p < .001$). For the main effect of academic domain, the mean percentage of broken DOI links for business ($M = 0.44$, $SD = 0.44$) was significantly greater than for the domain of health and medicine ($M = 0.33$, $SD = 0.39$, $p = .007$). For the main effect of academic domain, the mean of percent of broken DOI links for business ($M = 0.44$, $SD = 0.44$) was significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p < .001$). For the main effect of academic domain, the mean percentage of broken DOI links for health and medicine ($M = 0.33$, $SD = 0.39$) was significantly greater than for the domain of science/math/technologies ($M = 0.23$, $SD = 0.29$, $p = .03$). For the main effect of academic domain, the mean percentage of broken DOI links for science/math/technologies ($M = 0.23$, $SD = 0.29$) was significantly lesser than for the domain of social sciences ($M = 0.40$, $SD = 0.32$, $p < .001$).

This study represents one of the first to compare the percentage of broken links in general and DOI links specifically across disciplines, so no past research exists against which to compare the results. There is a significant difference in the percentage of broken DOI links among the disciplines addressed by this study. One possible factor contributing to these percentage differences involves the use of different style guides. As outlined in Chapter 2, style guides vary by discipline and, occasionally, by institution. Additionally, some study guides have changed more frequently than others, and some still contain outdated information in their most recent versions.

Science, math, and technology publications had the lowest percentage of broken DOI links. These disciplines most commonly use the *Chicago Manual of Style* as their chosen style guide. Chicago only allows the "https://doi.org/" format, which aligns with the International DOI Foundation recommendations. Science, math, and technology publications also have an alternative to DOI in ArXiv. ArXiv is unique because it does not just collect the metadata for a publication; it also acts as a repository for the publication itself. ArXiv can do this by limiting their publications to only those that are open access. The benefit of this method is that arXiv controls the publication's location, enabling them to ensure the reference remains up-to-date.

The style guide associated with a given discipline appears to play a role in the percentage of broken DOI links. Future researchers should further explore the connection between percentage of broken links and DOIs and style guide. Such a study should include an historical element to determine which style guide has the lowest percentage of broken links and DOIs over time. This could help inform the selection of style guides going forward. The use of DOIs in each discipline should trigger further investigation into the efficacy of DOI alternatives in combatting link rot.

Study Limitations

The limitations of this study centered around the availability of resources and the researcher's capacity to compute and codify results. First, the researcher had access to a limited number of scholarly publications, but the researcher also had limited capacity to compute and codify the results. Therefore, the researcher analyzed a limited sample size due to the computing power and time required to process a larger number of publications. Each publication required processing, codification, and review for errors or anomalies. The size of the dataset already delayed the completion of this study several times.

Another limitation involved the small number of ArXiv references that emerged in the dataset. ArXiv is a much less popular standard than DOI, but the original intention of this study was to include these links. In future studies, researchers should specifically focus on the availability of ArXiv references.

This study required the parsing of 2,500 articles containing 28,790 links. No tool existed to automate this job, especially considering the need to verify and analyze anomalies. As a result, the researcher developed a custom tool written in the Python programming language to accomplish the task. This process introduced limitations of programming experience and time. It took over a year to refine the Python tool to the point where it could successfully accomplish the task.

Implications for Future Practice

The biggest revelation of this study is that DOIs do not solve the problem of link rot. Developers created DOIs to address link rot, but they fell short of solving the problem and even created new ones in some instances. While no widely used or accepted substitute currently exists, the use of DOIs has proven just as unreliable as other links. Despite that fact, scholarly publishers will continue to lean on them until studies such as this one lead to a more effective treatment.

Importantly, this study shed light on the disparity between the way a style guide dictates a researcher cite sources and the way the persistent link system works. As a result, researchers should consider evaluating style guides for accuracy in this area. The bodies governing style guides should also invest heavily in efforts to stay current with the latest practices. The preferred solution would be for style guides to point researchers to those who maintain the persistent link system, such as the International DOI Foundation, for best practices.

The most recent research on link rot and content drift focused on systems that keep publications and their locations connected. While DOIs take metadata and redirect their links to content, systems like Memento, which Klein and Balakireva (2021) proposed, create a snapshot of a publication at the time of citation and direct link viewers to that archive rather than to a location that has potentially been changed or not maintained. The use of the Internet Archive's Wayback Machine would create a similar result. Unlike other systems, the Wayback Machine has a decades-long reputation for archival work and accessibility. Archiving a copy of all referenced publications is advisable going forward.

A reoccurring issue in this study involved links that directed users to content held behind a paywall. Paywalls present another barrier to source verification. Although business models require these measures for profitability, ways exist to avoid monetary losses. Best practice should involve pointing the link to the resource's publicly available abstract, not to the inaccessible location.

Additionally, researchers using a citation manager with automatic detection should review all fields that are not manually entered. This process will ensure all data and metadata have been entered and formatted correctly. Without a proper review, researchers may unknowingly compromise their own references.

Further, researchers should include the date of access in their references. This date clarifies what version of the article the researcher referenced. If archiving and snapshot systems like Memento and Wayback Machine become more common in scholarship, they will enable readers to navigate to the closest version of the article read by the researcher.

Another way to create more robust citations involves citing both the original link and a persistent link. Although no style guides currently advise this practice, it would enable navigation to the original content and a persistent version. This could help readers discover if the

publication has been retracted or changed. It would also allow users to revert to the original link if the persistent one is less maintained.

Lastly, those managing content and metadata should regularly check that all their links remain active and accurate. As early as 1998, internet researchers advised regularly running a link validator on every website (Krol & Zdonek, 2020). Scholarly publications maintaining the progression of knowledge should especially adopt this practice, but the benefits of regular link checks apply to all digital assets. Publishers and maintainers should conduct such checks on a regular basis and correct broken links. Not all researchers have the ability to conduct these checks, but this study led to the creation of the tools needed to facilitate the process.

Recommendations for Future Research

This study revealed a number of areas that would benefit from future research. The first recommendation involves the use of an already-established tool for gathering data. A large portion of time in this study was spent creating a tool to systematically produce a dataset. While the original project was created for use in the command line, the project expanded to include a web app that can be accessed at https://rottingresearch.org. Both tools can be found on the project's GitHub account at https://github.com/rottingresearch. Both projects are open-source and are free for use, modification, or distribution.

Another recommendation involves researching other systems and styles used to maintain access to resources. A number of other tools and services exist that aim to provide persistent access to resources. While DOI uses metadata to forward a user to a resource's location, other persistent link services also include the ability to archive the work itself. For a fee, Perma.cc archives resources at a certain point in time. The Berkman Klein Center for Internet and Society at Harvard Law School created and continues to support the Perma.cc project. Additionally, ArXiv asks users to upload their work with an open license so that ArXiv can store the resource

and provide consistent access over time. Even the Internet Archive's Wayback Machine has been used to archive resources and web pages. The Wayback Machine requires the visitor to know the date that the link was created so that it can return the result closest to the time of use.

In addition to changing practice, researchers should reproduce this study. Such research would help solidify trends and benchmarks. One of the most important outcomes of reproducing this study would be to set a generally accepted threshold value of link rot for this field of research. The issue of link rot appears far from resolved, so it could be helpful for future researchers to have this as a tool for their research.

Additionally, researchers should conduct more studies to understand better the relationship of link rot to subject areas. This study represents one of the first to compare the percentage of broken links and DOIs across disciplines, so no past research exists to compare the results for alignment. Future researchers should dig deeper to understand precisely why some disciplines have higher or lower percentages of broken links.

Another area of research could address the efficacy of different style guides in producing long-lasting links. Several issues arose from the fact that a few style guides changed how they handled DOI citations over time. This caused the production of unreachable links or even lost resources. These style guides have also changed how they cite other links as well. Chapter 2 addressed the shortening of URLs using http sites and ftp servers. A study tracking the historical reachability of links by style guide would inform scholars on which style guide might be best for their use. This study could also inform style guides authors on how to improve their guidelines.

A broader study should also be conducted to address the issue of reference rot. Reference rot includes link rot but adds the component of content drift, which occurs when resources change the content from its original appearance when cited. Although this study did not address this issue, it did show several resources had been changed from their original content over time.

Research into the cost of assuring that citations remain reachable would greatly benefit those exploring the field as well. Researchers should analyze the cost of link maintenance for publishers and institutions. This would help decision makers figure out how they can help to counteract link rot.

Finally, a study into how link rot affects an institution's ranking could be extremely helpful to those institutions looking to move up in the rankings. Several ranking systems exist for colleges, some of which influence funding, athletic conferences, and research. The most valuable study would focus on the Carnegie Classification system as it is most closely related to research. Numerous institutions have a goal to move up to become Research 1 institutions.

Conclusion

Link rot is an established phenomenon that affects everyone who uses the internet. The extent of link rot's influence on scholarly publications has been recently addressed by researchers looking at individual subjects. This study addressed link rot across disciplines. The study additionally compared the efficacy of DOIs and addressed what factors might influence the degradation of research over time. Consequently, this study will help inform the academic community and its leaders on how to address this issue.

A significant difference in the percentage of broken links existed among academic disciplines, but further research is required to identify what influences these differences. Citations and style guides that deviate from best practices make it hard to fully address the issue.

DOIs proved not to be the solution developers promised, but tools created from research like Memento, Perma.cc, ArXiv, and the Wayback Machine offer hope. The prevailing suggestion is that a snapshot or archive may represent the best method for solving the issue of link rot. Further research is required to find a suitable solution, but institutions and researchers can make changes to their practice to mitigate the damage.

References

Abt, H. A. (2018). What fraction of papers in astronomy and physics are not cited in 40 years? *Publications of the Astronomical Society of the Pacific, 130*(989), 1–4.

Adams, K. A., & Lawrence, E. K. (2018). *Research Methods, Statistics, and Applications* (2nd edition). SAGE Publications, Inc.

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, *9*(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.).

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.).

American University Librarians. (2022). *Subject guides: Citation style guide: Which style should I use? What style should I use?* https://subjectguides.library.american.edu/c.php?g=175008&p=1154150

The Associated Press. (2020). *The Associated Press stylebook* (55th ed.). Basic Books.

Bradshaw, C. J. A., Chalker, J. M., Crabtree, S. A., Eijkelkamp, B. A., Long, J. A., Smith, J. R., Trinajstic, K., & Weisbecker, V. (2021). A fairer way to compare researchers at any career stage and in any discipline using open-access citation data. *PLoS ONE, 16*(9), 1–15. https://doi.org/10.1371/journal.pone.0257141

Broadus, R. N. (1983). An investigation of the validity of bibliographic citations. *Journal of the American Society for Information Science*, *34*(2), 132–135. https://doi.org/10.1002/asi.4630340206

Callister, P. D. (2021). Perma.cc and web archival dissonance with copyright law. *Legal Reference Services Quarterly*, *40*(1), 1–57. https://doi.org/10.1080/0270319X.2021.1886785

Chrome 86: Improved focus highlighting, WebHID, and more. (2020). *Chromium Blog*. Google. https://blog.chromium.org/2020/09/chrome-86-improved-focus-highlighting.html

Colavizza, G., Romanello, M., & Kaplan, F. (2018). The references of references: A method to enrich humanities library catalogs with citation data. *International Journal on Digital Libraries*, *19*(2/3), 151–161. https://doi.org/10.1007/s00799-017-0210-1

Edmonds, W. A., & Kennedy, T. D. (2016). *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods* (Second edition). SAGE Publications, Inc.

Electronic Frontier Foundation. (2022, April 15). *Let's Encrypt wins Levchin Prize for work on internet security*. https://www.eff.org/press/releases/lets-encrypt-wins-levchin-prize-work-internet-security

Elsevier fact sheet. (2018, January 7). MIT Libraries. https://libraries.mit.edu/scholarly/publishing/elsevier-fact-sheet/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics*. SAGE Publications, Inc.

Fraenkel, J., Wallen, N., & Hyun, H. (2018). *How to Design and Evaluate Research in Education* (10th edition). McGraw Hill.

Genzinger, P., & Wills, D. (2017). Giving credit: How well do librarians cite and quote their sources? *Reference & User Services Quarterly*, *57*(1), 30–41.

George, D., & Mallery, P. (2019). *IBM SPSS Statistics 26 step by step: A simple guide and reference* (16th ed.). Routledge. https://doi.org/10.4324/9780429056765

Gerow, A., Hu, Y., Boyd-Graber, J., Blei, D. M., & Evans, J. A. (2018). Measuring discursive influence across scholarship. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3308–3313.

Hassan, N. R., & Serenko, A. (2019). Patterns of citations for the growth of knowledge: A Foucauldian perspective. *Journal of Documentation*, *75*(3), 593–611. https://doi.org/10.1108/JD-08-2018-0125

Hendricks, G. (2021). *About us*. Crossref. https://www.crossref.org/about/

Homenda, N. (2021). Persistent URLs and citations offered for digital objects by digital libraries. Information *Technology & Libraries*, *40*(2), 1–12. https://doi.org/10.6017/ital.v40i2.12987

The HTTPS-only standard. (2015). Chief Information Officers Council. https://https.cio.gov/

International DOI Foundation. (2016, October 17). *Table of contents*. DOI. https://www.doi.org/doi_handbook/TOC.html

International DOI Foundation. (2020, June 7). *What are persistent identifiers (PIDs)?* ORCID. https://support.orcid.org/hc/en-us/articles/360006971013

Jackson, L. J. (2013, December 1). "Link rot" is degrading legal research and case cites. *ABA Journal*. https://www.abajournal.com/magazine/article/link_rot_is_degrading_legal_research_and_case_cites/

Jomaa, N. J., & Bidin, S. J. (2019). Variations in the citation use and perceptions in writing the literature review by EFL postgraduates. *Eurasian Journal of Applied Linguistics*, *5*(3), 441–460.

Jones, B. F. (2021). The rise of research teams: Benefits and costs in economics. *Journal of Economic Perspectives*, *35*(2), 191–216. https://doi.org/10.1257/jep.35.2.191

Kafes, H. (2017). Citation practices among novice and expert academic writers. *Education and Science*, *42*(192), 442–462. https://www.proquest.com/docview/1966351006

Kaufman, R., & Campana, A. (2019). OER: Lessons from the field. *Insights the UKSG Journal*, *32*, 15. https://doi.org/10.1629/uksg.464

Klein, M., & Balakireva, L. (2021). An extended analysis of the persistence of persistent identifiers of the scholarly web. *International Journal on Digital Libraries*, *23*, 5–17. https://doi.org/10.1007/s00799-021-00315-w

Krol, K., & Zdonek, D. (2020). Peculiarity of the bit rot and link rot phenomena. *Global Knowledge, Memory and Communication*, *69*(1/2), 20–37. http://dx.doi.org/10.1108/GKMC-06-2019-0067

Lerner, N., & Oddis, K. (2017). The social lives of citations: how and why *Writing Center Journal* authors cite sources. *The Writing Center Journal*, *36*(2), 235–262.

Milewski, S. D., Williamson, J. M., & Trott, B. (2017). Developing a reflective practice template for citation management software instruction. *Reference & User Services Quarterly*, *57*(1), 6–11.

Mills, D., Branford, A., Inouye, K., Robinson, N., & Kingori, P. (2021). "Fake" journals and the fragility of authenticity: Citation indexes, "predatory" publishing, and the African research ecosystem. *Journal of African Cultural Studies*, *33*(3), 276–296. https://doi.org/10.1080/13696815.2020.1864304

Modern Language Association. (2016). *Modern Language Association handbook* (8th ed.).

Modern Language Association. (2021). *Modern Language Association handbook* (9th ed.).

Penders, B. (2018). Ten simple rules for responsible referencing. *PLoS Computational Biology*, *14*(4), 1–6. https://doi.org/10.1371/journal.pcbi.1006036

Perkel, J. M. (2015). The trouble with reference rot. *Nature*, *521*(7550), 111–112. https://doi.org/10.1038/521111a

Rivkin, A. (2020). Manuscript referencing errors and their impact on shaping current evidence. *American Journal of Pharmaceutical Education*, *84*(7), 877–880. https://doi.org/10.5688/ajpe7846

Robinson, D. C., Hand, J. A., Madsen, M. B., & Mckelvey, K. R. (2018). The Dat Project, an open and decentralized research data tool. *Scientific Data*, *5*, 180221. http://dx.doi.org/10.1038/sdata.2018.221

Rozell, D. J. (2020). *Dangerous Science*. Ubiquity Press. https://www.jstor.org/stable/j.ctv11cvx39.5

Rubbo, P., Pilatti, L. A., & Picinin, C. T. (2019). Citation of retracted articles in engineering: A study of the web of science database. *Ethics & Behavior*, *29*(8), 661–679. https://doi.org/10.1080/10508422.2018.1559064

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Scholarly Communication and Open Publishing. (2020, June 26). SCOOP: *From DOIs to ORCIDs—How persistent identifiers work (and how to make them work for you)*. Atla. https://www.atla.com/blog/scoop-how-persistent-identifiers-work/

Shemberko, L. V., & Shershova, A. V. (2018). Normative and technological problems of bibliographic control over information resources on social sciences and humanities. *Scientific and Technical Information Processing*, *45*(1), 55–62. https://doi.org/10.3103/s0147688218010100

Citation context analysis. In B. Dervin & M. Voigt (Eds.), Progress in communication sciences (Vol. 3, pp. 287-310)

Turabian, K. L., Booth, W. C., Colomb, G. G., Williams, J. M., Bizup, J., & FitzGerald, W. T. (2018). *A manual for writers of research papers, theses, and dissertations: Chicago Style for students and researchers* (9th edition). The University of Chicago Press.

University of Chicago Press Staff. (2003). *The Chicago manual of style* (15th ed.). University of Chicago Press.

University of Chicago Press Staff. (2017). *The Chicago manual of style* (17th ed.). University of Chicago Press.

Urhan, T. K., Rempel, H. G., Meunier-Goddik, L., & Penner, M. H. (2018). Information retrieval in food science research: A bibliographic database analysis. *Journal of Food Science, 83*(12), 2912–2922. https://doi.org/10.1111/1750-3841.14388

West, R., Stenius, K., & Kettunen, T. (2017). Use and abuse of citations. In K. Stenius, T. F. Babor, R. Pates, M. Miovsky, J. O'Reilly, & P. Candon (Eds.), *Publishing addiction science* (pp. 191–206). Ubiquity Press. https://www.jstor.org/stable/j.ctv3t5qxw.16

Zhao, S. X., & Qiao, L. (2017). Subject ranking of universities. *Current Science*, *113*(7), 1214–1215.

Zittrain, J. L., Albert, K., & Lessig, L. (2013). *Perma: Scoping and addressing the problem of link and reference rot in legal citations* (SSRN Scholarly Paper ID 2329161). Social Science Research Network. https://doi.org/10.2139/ssrn.2329161

Appendix

Pilot Study Results

Publication #	Total Links	Broken Links	Valid Links	301	302	400	403	404	405	500	503	522	DOI	Emails	% Valid
1	21	7	14	0	0	0	0	1	0	0	0	0	6	1	67%
2	52	11	41	0	1	0	1	8	0	0	0	0	1	1	79%
3	13	2	11	0	0	0	0	2	0	0	0	0	0	1	85%
4	208	174	34	0	1	1	0	12	0	0	0	0	160	1	16%
5	35	17	18	0	1	0	0	5	0	0	0	0	11	1	51%
6	66	7	59	0	2	0	0	4	0	0	0	0	1	1	89%
7	19	7	12	0	0	0	0	1	0	0	0	0	6	1	63%
8	23	4	19	0	0	0	0	4	0	0	0	0	0	1	83%
9	39	7	32	0	0	0	0	6	0	0	0	1	0	1	82%
10	8	0	8	0	0	0	0	0	0	0	0	0	0	1	100%
11	16	2	14	0	0	0	1	0	0	0	0	0	1	1	88%
12	46	13	33	0	0	0	0	10	0	0	0	0	3	1	72%
13	40	17	23	0	0	0	1	2	0	0	0	0	14	1	58%
14	35	12	23	0	0	0	0	2	0	0	0	0	10	1	66%

15	28	1	27	0	0	0	0	1	0	0	0	0	0	1	96%
16	16	3	13	0	0	0	1	1	0	0	0	0	1	1	81%
17	51	9	42	0	1	0	0	8	0	0	0	0	0	1	82%
18	68	32	36	0	0	0	1	16	0	0	0	0	15	1	53%
19	59	32	27	0	0	1	0	16	0	2	0	0	13	1	46%
20	41	25	16	0	0	0	0	2	0	0	0	0	23	1	39%
21	59	13	46	0	0	0	1	6	0	0	1	0	5	2	78%
22	16	3	13	0	0	0	0	1	0	0	0	0	2	1	81%
23	46	32	14	0	0	0	0	4	0	0	0	0	28	6	30%
24	68	10	58	0	0	4	0	2	0	0	0	0	4	1	85%
25	83	30	53	0	0	0	2	9	0	0	0	0	19	0	64%
26	73	50	23	0	0	0	0	0	0	0	0	0	50	1	32%
27	15	3	12	0	0	0	0	0	0	0	0	0	3	1	80%
28	60	33	27	0	0	0	0	5	0	0	0	0	28	1	45%
29	29	21	8	0	0	0	1	4	0	4	1	0	11	1	28%
30	145	98	47	0	0	0	0	9	0	0	0	0	89	1	32%
31	62	50	12	0	0	0	0	3	0	0	0	0	47	1	19%
32	45	13	32	0	0	0	0	0	1	0	0	0	12	3	71%

33	114	49	65	1	0	0	2	5	0	0	0	0	41	2	57%
34	43	1	42	0	0	0	0	1	0	0	0	0	0	1	98%
35	107	34	73	0	0	0	0	10	0	0	0	0	24	1	68%
36	64	14	50	0	0	0	0	3	0	0	0	0	11	1	78%
37	30	15	15	0	0	0	0	1	0	0	0	0	14	1	50%
38	45	20	25	0	0	0	1	5	0	0	0	0	14	1	56%
39	66	7	59	0	0	0	0	2	0	0	0	0	5	1	89%
40	22	8	14	0	0	0	4	0	0	1	0	0	3	4	64%
41	14	1	13	0	0	0	0	1	0	0	0	0	0	1	93%

Mean Avg.	73%	
Average	58%	
Total Links	2090	
Broken Links	887	**% of Broken Links**
DOI	675	76%
301	1	0%
302	6	1%

400	6	1%
403	16	2%
404	172	19%
405	1	0%
500	7	1%
503	2	0%
522	1	0%