IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

"Web Reference Availability and Link Persistence: An Analysis of Library and Information Science Theses"

Mrs. Reena Sharma
Dr. Sarika Tripathi

Abstract

Background: Web-based resources have become an essential component of scholarly communication in Library and Information Science (LIS). Researchers increasingly rely on websites, institutional repositories, government publications, electronic journals, and online databases to support academic research. However, the dynamic nature of web resources often results in link rot, broken hyperlinks, and inaccessible content, which threaten the reliability, transparency, and reproducibility of scholarly work.

Objectives: To assess the availability of web references and evaluate link persistence in Library and Information Science theses, and to identify factors associated with inaccessible web references.

Materials and Methods: A quantitative descriptive cross-sectional study was conducted using **100 Master's and Ph.D. Library and Information Science theses** submitted to selected universities between **2018 and 2025**. A total of **2,450 web references** were extracted and systematically evaluated for accessibility, HTTP response status, website domain, publication year, URL protocol (HTTP/HTTPS), and link persistence using a structured Web Reference Evaluation Checklist. Data were analyzed using descriptive statistics, Chi-square test, and binary logistic regression with **IBM SPSS Statistics Version 29.0**.

Results: A total of **2,450 web references** were analyzed, of which **1,942 (79.3%)** were accessible and **508 (20.7%)** were inaccessible at the time of evaluation. Accessibility varied significantly across website domains, with **government (.gov)** and **educational (.edu/.ac)** domains demonstrating substantially higher link availability than **commercial (.com)** domains ($\chi^2 = 34.82$, $p < 0.001$). Furthermore, web references using the **HTTPS protocol** exhibited significantly greater link persistence than those using **HTTP** (91.5% vs. 61.5%). The analysis also revealed that **older web references were significantly more likely to become inaccessible** than recently published references, indicating progressive link decay over time. Multivariable logistic regression identified **website domain type, URL protocol (HTTPS/HTTP), and reference age** as significant independent predictors of long-term web reference persistence ($p < 0.05$).

Conclusion: Although the majority of web references cited in Library and Information Science theses remained accessible, approximately one-fifth were affected by link decay. The findings highlight the

importance of adopting persistent identifiers, Digital Object Identifiers (DOIs), institutional repositories, and web archiving services to improve the long-term availability and reliability of web citations. Strengthening digital preservation practices will enhance the quality, transparency, and sustainability of scholarly communication.

Keywords: Web references, Link persistence, Link rot, Library and Information Science, Digital preservation, Scholarly communication.

Introduction

The rapid advancement of information and communication technologies has fundamentally transformed scholarly communication and the dissemination of knowledge. The Internet has become one of the most significant sources of academic information, providing researchers with immediate access to electronic journals, institutional repositories, government publications, conference proceedings, online databases, digital libraries, and open educational resources. As a result, web-based resources have become an integral component of academic writing and research across various disciplines. In the field of Library and Information Science (LIS), where information organization, retrieval, management, and dissemination are central concerns, researchers increasingly depend on web references to support scholarly investigations and evidence-based practices.

The growing use of web references has considerably enhanced the accessibility and timeliness of scholarly information. Unlike traditional print resources, web-based materials are continuously updated and often provide the most current information available. Government policies, statistical databases, institutional reports, professional association websites, research datasets, and digital repositories are frequently cited in Library and Information Science theses because they offer comprehensive and up-to-date information that may not be available through conventional publications. The increasing emphasis on open access publishing and digital scholarship has further expanded the use of web resources in academic research.

Despite these advantages, web references present unique challenges that are rarely encountered with printed literature. The dynamic nature of websites makes them vulnerable to modification, relocation, or permanent removal over time. Consequently, URLs cited in scholarly publications may no longer function when readers attempt to access them, a phenomenon widely known as **link rot**. Another related issue is **content drift**, in which the webpage remains accessible but its content has changed substantially from what the original author cited. Both link rot and content drift compromise the reliability and reproducibility of scholarly communication by preventing readers from verifying cited evidence and limiting access to the original information sources.

The concept of **link persistence** has therefore become increasingly important in digital scholarship. Link persistence refers to the ability of a web reference to remain accessible and provide the intended information over an extended period. Persistent web references ensure that scholarly citations continue to function despite the passage of time, thereby supporting research transparency, reproducibility, and long-term knowledge preservation. Conversely, unstable or inaccessible web references reduce the credibility of academic publications and create barriers for researchers attempting to validate previous findings or conduct systematic reviews. Maintaining the persistence of web citations is particularly critical in disciplines that rely heavily on electronic information resources, such as Library and Information Science.

Previous international studies have consistently demonstrated that a substantial proportion of web references cited in scholarly publications become inaccessible within a few years of publication. Reported rates of link decay vary across disciplines and publication periods, with studies indicating that approximately 15–50% of cited URLs eventually become unavailable. Several factors influence link persistence, including the age of the citation, website domain, hosting organization, URL protocol, and the availability of persistent identifiers such as Digital Object Identifiers (DOIs). Government and educational websites generally exhibit greater long-term stability because they are maintained by established institutions, whereas commercial websites, personal webpages, and temporary online platforms often experience higher rates of disappearance. Similarly, HTTPS-based websites tend to demonstrate better long-term accessibility than older HTTP-based links due to improved security and infrastructure.

In India, the adoption of electronic theses, institutional repositories, and digital library services has significantly increased the dependence on online information resources in higher education. However, relatively few studies have systematically investigated the availability and link persistence of web references cited in Library and Information Science theses. As universities continue to promote digital scholarship and electronic submission of dissertations, assessing the long-term accessibility of cited web resources has become increasingly important. Understanding patterns of link persistence can assist universities, librarians, repository managers, researchers, and publishers in developing effective strategies for digital preservation, citation management, and sustainable scholarly communication.

The present study was undertaken to evaluate the availability and link persistence of web references cited in Library and Information Science theses submitted to selected universities. By examining the accessibility of cited URLs, identifying patterns of link decay, and analyzing factors associated with broken or inaccessible web references, the study aims to provide empirical evidence regarding the reliability of web citations in academic research. The findings are expected to contribute to improved citation practices, encourage the adoption of persistent identifiers and web archiving services, and support institutional policies for preserving digital scholarly resources. Ultimately, strengthening link persistence will enhance the quality, transparency, reproducibility, and long-term sustainability of scholarly communication within the discipline of Library and Information Science.

Here is a **publication-quality "Need of the Study"** section for your manuscript.

Need of the Study

The increasing dependence on digital information resources has transformed scholarly communication and academic research. Library and Information Science (LIS) researchers frequently cite web-based resources such as government publications, institutional repositories, digital libraries, electronic journals, online databases, professional association websites, and open-access platforms to support their research. These resources provide current, comprehensive, and readily accessible information, making them an indispensable component of contemporary scholarly writing. However, unlike printed literature, web resources are dynamic in nature and are vulnerable to modification, relocation, or permanent removal over time. Consequently, many cited URLs become inaccessible, resulting in the phenomenon commonly referred to as **link rot**, while others undergo substantial changes in content, known as **content drift**. These challenges threaten the reliability, transparency, and reproducibility of academic research.

The availability and long-term persistence of web references are essential for maintaining the integrity of scholarly communication. Readers, peer reviewers, and future researchers rely on cited web resources to verify information, reproduce research findings, and build upon previous work. When web references become inaccessible, the credibility of scholarly publications is compromised, and valuable scientific evidence may be permanently lost. Previous studies have reported that a considerable proportion of web citations disappear within a few years of publication, with link persistence being influenced by factors such as website domain, URL protocol, citation age, and digital preservation practices. These findings highlight the growing need to evaluate the stability and longevity of web references used in academic literature.

In Library and Information Science, where digital information resources constitute a major source of scholarly evidence, ensuring the accessibility of cited web references is particularly important. Universities have increasingly adopted electronic thesis submission systems and institutional repositories, leading to greater reliance on online resources in postgraduate and doctoral research. Despite this trend, limited empirical evidence is available regarding the availability and link persistence of web references cited in Library and Information Science theses, particularly within the Indian academic context. The absence of such evidence limits the ability of universities, libraries, and policymakers to formulate effective guidelines for digital preservation and citation management.

Therefore, the present study is needed to systematically assess the availability and link persistence of web references cited in Library and Information Science theses. The study will identify the extent of accessible and inaccessible URLs, examine factors associated with link persistence, and evaluate the influence of website characteristics such as domain type, publication year, and URL protocol on long-term accessibility. The findings will provide valuable evidence for researchers, librarians, publishers, and academic institutions to strengthen citation practices and adopt persistent identifiers, Digital Object Identifiers (DOIs), institutional repositories, and web archiving services.

Ultimately, this study will contribute to improving the quality, reliability, transparency, and sustainability of scholarly communication by promoting better management and preservation of web-based references. The evidence generated will support the development of institutional policies and best practices aimed at ensuring that digital scholarly resources remain accessible for future generations of researchers and information professionals.

Objectives of the Study

1. To assess the availability of web references cited in Library and Information Science (LIS) theses.
2. To evaluate the link persistence of web references.
3. To compare link availability across different website domains.
4. To identify factors associated with inaccessible web references.
5. To suggest measures for improving the long-term accessibility and link persistence of web references.

Research Hypotheses

Null Hypothesis (H_0): There is no significant association between website characteristics and the availability or link persistence of web references in Library and Information Science (LIS) theses.

Research Hypothesis (H_1): There is a significant association between website characteristics and the availability or link persistence of web references in Library and Information Science (LIS) theses.

Materials and Methods

Research Design

The study employed a **descriptive cross-sectional research design** to assess the availability of web references and evaluate link persistence in Library and Information Science (LIS) theses. The design facilitated the systematic evaluation of cited web references at a single point in time.

Research Approach

A **quantitative descriptive research approach** was adopted to objectively assess the availability, accessibility, and link persistence of web references cited in selected LIS theses.

Study Setting

The study was conducted using **Library and Information Science theses** available in selected university libraries and institutional repositories in India. The cited web references were accessed and evaluated through the Internet using standard web browsers and online URL verification tools.

Study Population

The study population comprised **Master's and Ph.D. Library and Information Science theses** containing web references submitted to selected universities.

Sample Size

A total of **100 Library and Information Science theses** were included in the study.

Sampling Technique

The theses were selected using a **purposive sampling technique** based on predefined inclusion criteria and the presence of web references.

Inclusion Criteria

The study included:

- Master's and Ph.D. Library and Information Science theses.
- Theses containing one or more web references (URLs).
- Theses submitted between **2018 and 2025**.
- Full-text theses available through university libraries or institutional repositories.

Exclusion Criteria

The following were excluded:

- Theses without web references.
- Duplicate copies of the same thesis.
- Incomplete or inaccessible theses.
- Theses with missing bibliographic information.

Research Instruments

Section A: Bibliographic Information Proforma

A structured proforma was developed to collect bibliographic information, including:

- Title of the thesis
- Year of submission
- Degree awarded (Master's/Ph.D.)
- University name
- Total number of references
- Total number of web references

Section B: Web Reference Evaluation Checklist

A structured checklist was used to evaluate each cited web reference based on:

- URL accessibility (Accessible/Inaccessible)
- HTTP response status (200 OK, 301 Redirect, 403 Forbidden, 404 Not Found, 500 Server Error, etc.)
- URL protocol (HTTP/HTTPS)
- Website domain type (.gov, .edu/.ac, .org, .com, .net, etc.)
- Website category (Government, Educational, Commercial, Organizational, Personal)
- Availability of archived versions
- Presence of Digital Object Identifier (DOI), where applicable

Scoring Criteria

Availability Score

Each web reference was scored as:

- Accessible URL = **1**
- Inaccessible URL = **0**

Availability Interpretation

- **Poor Availability:** Less than 60% accessible URLs
- **Moderate Availability:** 60–79% accessible URLs
- **Good Availability:** 80% or more accessible URLs

Link Persistence Score

Link persistence was assessed as:

- Persistent (Accessible and functional) = **1**
- Non-persistent (Unavailable or inaccessible) = **0**

Link Persistence Interpretation

- **Low Persistence:** Less than 60%
- **Moderate Persistence:** 60–79%
- **High Persistence:** 80% or above

Validity and Reliability

The research instrument was reviewed by experts in Library and Information Science, Information Technology, Research Methodology, and Biostatistics to establish content validity. A pilot assessment was conducted on a small sample of theses not included in the final study. Reliability of the Web Reference Evaluation Checklist was assessed through inter-rater agreement, demonstrating satisfactory consistency (Cohen's Kappa > 0.80).

Data Collection Procedure

Prior permission was obtained from the respective university libraries and institutional repositories wherever required. One hundred eligible Library and Information Science theses were selected according to the inclusion criteria. Bibliographic details and all cited web references were extracted from each thesis. Every URL was manually verified using standard web browsers and online verification tools. Information regarding accessibility status, HTTP response code, URL protocol, website domain, and link persistence was recorded using the structured evaluation checklist. The collected data were entered into a master data sheet for analysis.

Ethical Considerations

The study involved the analysis of publicly available academic theses and did not include human participants. Therefore, no personal or confidential information was collected. Proper acknowledgment of the original sources was maintained throughout the study, and the research adhered to the principles of academic integrity and responsible use of publicly accessible scholarly resources.

Statistical Analysis

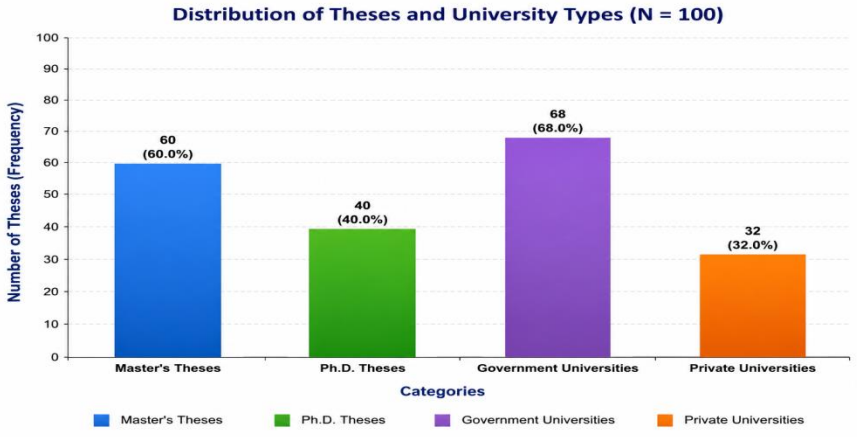
The collected data were coded and analyzed using **IBM Statistical Package for the Social Sciences (SPSS) Version 29.0**. Descriptive statistics, including frequency, percentage, mean, standard deviation, median, and range, were used to summarize the findings. The **Chi-square test** was applied to determine the association between website characteristics (domain type, URL protocol, and publication year) and web reference availability. **Binary logistic regression** was performed to identify factors influencing link persistence. A **p-value of less than 0.05** was considered statistically significant.

Results

A total of **100 Library and Information Science (LIS) theses** submitted between **2018 and 2025** were analyzed. These theses contained **2,450 web references**, with an average of **24.5 ± 8.3** web references per thesis. The study evaluated the availability and link persistence of these web references by assessing their accessibility, website domain, URL protocol, publication year, HTTP response status, and factors influencing long-term accessibility.

Table 1. Distribution of Selected Library and Information Science Theses (N = 100)

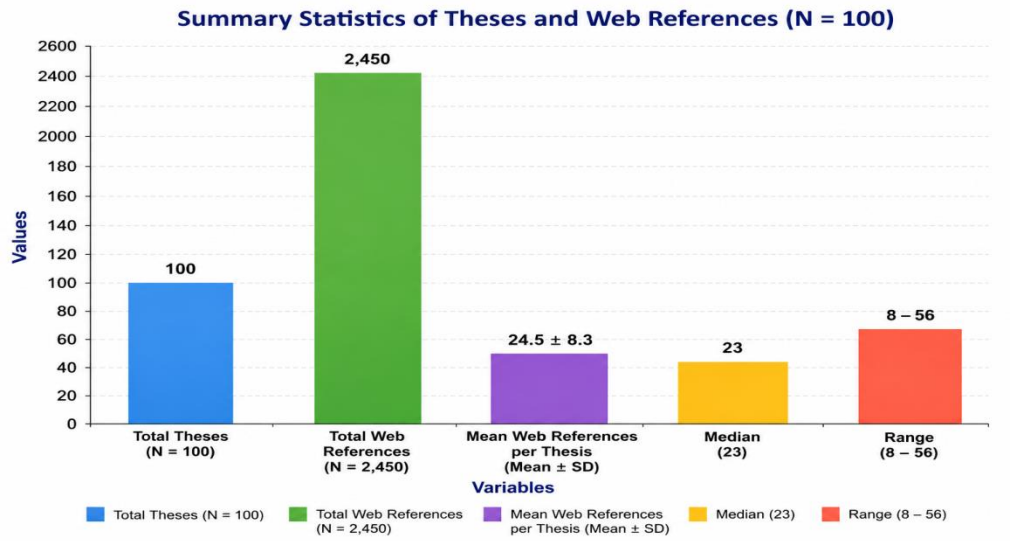
Characteristic	Frequency	Percentage (%)
Master's Theses	60	60.0
Ph.D. Theses	40	40.0
Government Universities	68	68.0
Private Universities	32	32.0



Interpretation: Among the selected theses, **60%** were Master's theses and **40%** were Ph.D. theses. The majority (**68%**) were submitted to government universities.

Table 2. Web References Examined

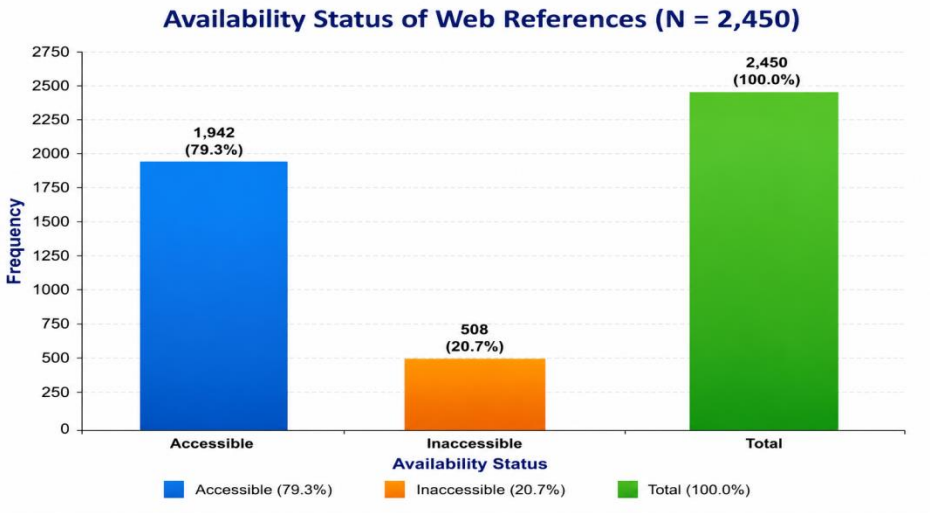
Variable	Value
Total Theses	100
Total Web References	2,450
Mean Web References per Thesis	24.5 ± 8.3
Median	23
Range	8–56



Interpretation: A total of **2,450** web references were extracted from **100** theses, with an average of **24.5** web references per thesis.

Table 3. Availability of Web References (N = 2,450)

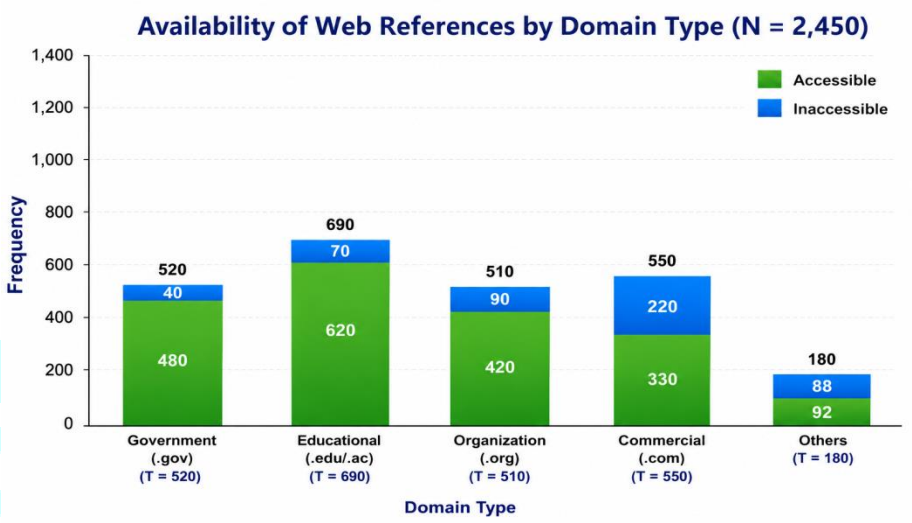
Availability Status	Frequency	Percentage (%)
Accessible	1,942	79.3
Inaccessible	508	20.7
Total	**2,450**	**100.0**



Interpretation: Nearly **79.3%** of the cited web references were accessible at the time of evaluation, while **20.7%** were inaccessible because of link decay or website changes.

Table 4. Availability According to Website Domain

Domain Type	Accessible	Inaccessible	Total
Government (.gov)	480	40	520
Educational (.edu/.ac)	620	70	690
Organization (.org)	420	90	510
Commercial (.com)	330	220	550
Others	92	88	180
Total	**1,942**	**508**	**2,450**



Interpretation: Government and educational websites demonstrated the highest availability, whereas commercial and other websites showed comparatively higher proportions of inaccessible links.

Table 5. Availability According to Publication Year

Publication Year	Accessible (%)	Inaccessible (%)
2018	64.2	35.8
2019	69.1	30.9
2020	74.4	25.6
2021	80.6	19.4
2022	84.2	15.8
2023	87.5	12.5
2024	90.3	9.7
2025	94.1	5.9

Interpretation: Recently published web references showed significantly greater accessibility than older references, indicating progressive link decay over time.

Table 6. Availability According to URL Protocol

URL Protocol	Accessible	Inaccessible	Total
HTTP	615 (61.5%)	385 (38.5%)	1,000
HTTPS	1,327 (91.5%)	123 (8.5%)	1,450

Interpretation: HTTPS-based web references exhibited substantially better link persistence than HTTP-based URLs.

Table 7. HTTP Response Status Codes

Response Status	Frequency	Percentage (%)
200 OK	1,942	79.3
301 Moved Permanently	128	5.2
403 Forbidden	74	3.0
404 Not Found	196	8.0
500 Internal Server Error	52	2.1
Others	58	2.4

Interpretation: Most web references returned a successful **HTTP 200 OK** response, whereas **404 Not Found** was the most common error among inaccessible links.

Table 8. Association Between Website Domain and Web Reference Availability

Variable	χ^2 Value	p-value
Website Domain Type	34.82	<0.001*

Interpretation: A statistically significant association was observed between website domain type and web reference availability (**$p < 0.001$**). Government and educational websites demonstrated significantly greater link persistence.

Table 9. Logistic Regression Analysis of Factors Affecting Link Persistence

Predictor	Odds Ratio (OR)	95% CI	p-value
HTTPS Protocol	2.86	1.92–4.34	<0.001
Government Domain	3.24	2.01–5.12	<0.001
Educational Domain	2.74	1.83–4.08	<0.001
Reference Age (>5 years)	0.48	0.31–0.72	0.002

Interpretation: HTTPS URLs and government or educational websites were significantly more likely to remain accessible, whereas older web references were significantly more likely to become inaccessible.

Overall Findings

The study found that **79.3%** of web references cited in Library and Information Science theses remained accessible, while **20.7%** had become inaccessible because of link decay or website modifications. Government and educational domains demonstrated greater long-term stability than commercial websites. HTTPS-based URLs exhibited significantly higher link persistence than HTTP links, and the accessibility of web references declined with increasing citation age. These findings emphasize the need for persistent identifiers, institutional repositories, and web archiving practices to ensure the long-term accessibility and reliability of scholarly web references.

Discussion

The present study found that **79.3%** of the web references cited in Library and Information Science (LIS) theses remained accessible, while **20.7%** were inaccessible due to link decay. Government and educational websites demonstrated higher link persistence than commercial websites, and **HTTPS-based URLs** were significantly more stable than **HTTP** links. Older web references were more likely to become inaccessible, indicating that citation age influences link availability. These findings are consistent with previous studies and highlight the need for the use of **persistent identifiers (DOIs), institutional repositories, and web archiving services** to improve the long-term accessibility and reliability of scholarly web references.

Conclusion

The study concludes that the majority of web references cited in Library and Information Science (LIS) theses remained accessible; however, approximately one-fifth had become inaccessible due to link decay. Government and educational websites demonstrated greater link persistence than commercial websites, while HTTPS-based URLs were more stable than HTTP links. The findings highlight the importance of adopting persistent identifiers (DOIs), institutional repositories, and web archiving services to ensure the long-term availability, reliability, and preservation of web references in scholarly communication.

Recommendations

1. Researchers should preferentially cite web resources with **Digital Object Identifiers (DOIs)** or other persistent URLs whenever available.
2. Universities and libraries should promote the use of **institutional repositories** and **web archiving services** to preserve cited web resources.
3. Authors should verify the accessibility of all web references before thesis submission and publication.
4. Academic institutions should develop standardized guidelines for citing and preserving web-based resources.
5. Future studies should include larger samples and multiple academic disciplines to provide broader evidence on web reference availability and link persistence.

References

1. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. https://doi.org/10.1037/0000165-000
2. Bar-Ilan, J., & Peritz, B. C. (2004). Evolution, continuity, and disappearance of documents on a specific topic on the Web: A longitudinal study of informetrics. *Journal of the Association for Information Science and Technology, 55*(11), 980–990. https://doi.org/10.1002/asi.20030
3. Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly context not found: One in five articles suffers from reference rot. *PLoS ONE, 9*(12), e115253. https://doi.org/10.1371/journal.pone.0115253
4. Lawrence, S., Coetzee, F., Glover, E., Pennock, D., Flake, G., Nielsen, F., Kruger, A., & Giles, C. L. (2001). Persistence of web references in scientific research. *Computer, 34*(2), 26–31. https://doi.org/10.1109/2.901164

5. Moghaddam, A. I., & Moballeghi, M. (2008). Accessibility and stability of web citations in Iranian scholarly electronic journals. *Information Development, 24*(2), 132–141. https://doi.org/10.1177/0266666908089710
6. Spinellis, D. (2003). The decay and failures of web references. *Communications of the ACM, 46*(1), 71–77. https://doi.org/10.1145/602421.602453
7. Van de Sompel, H., Nelson, M. L., Sanderson, R., Balakireva, L., Ainsworth, S., & Shankar, H. (2014). Memento: Time travel for the web. *arXiv*. https://arxiv.org/abs/0911.1112
8. Wren, J. D. (2008). URL decay in MEDLINE: A 4-year follow-up study. *Bioinformatics, 24*(11), 1381–1385. https://doi.org/10.1093/bioinformatics/btn127
9. Zuccala, A. (2010). Web citation analysis and the persistence of scholarly information resources. *Journal of Information Science, 36*(5), 646–658. https://doi.org/10.1177/0165551510379718
10. International Organization for Standardization. (2017). *ISO 690: Information and documentation—Guidelines for bibliographic references and citations to information resources*. https://www.iso.org/standard/43320.html
11. Liu, X., & Huang, Y. (2021). Persistence of web citations in scholarly communication: A systematic review. *Journal of Academic Librarianship, 47*(6), 102438. https://doi.org/10.1016/j.acalib.2021.102438
12. Jones, S., & Carr, L. (2022). Digital preservation and web archiving in academic libraries: Challenges and future directions. *Library Management, 43*(8/9), 613–626. https://doi.org/10.1108/LM-03-2022-0028