



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

“AVAILABILITY AND PERSISTENCY OF WEB REFERENCES: A STUDY OF LIBRARY AND INFORMATION SCIENCE THESES”

Mrs. Reena Sharma

Dr. Sarika Tripathi

Abstract

Background: Web resources have become an integral part of scholarly communication in Library and Information Science (LIS). However, the long-term availability and persistency of web references remain a major concern because websites frequently change, move, or disappear over time. Link rot and content drift reduce the reliability of academic literature and hinder future researchers from accessing cited sources.

Objectives: To assess the availability and persistency of web references cited in Library and Information Science theses and identify factors associated with inaccessible web resources.

Methods: A descriptive cross-sectional study was conducted on **100 postgraduate and doctoral Library and Information Science theses** submitted to selected universities between 2018 and 2025. A total of **2,450 web references** were extracted and manually verified. Each URL was examined for accessibility, HTTP response status, domain type, publication year, protocol (HTTP/HTTPS), and content availability. Data were analyzed using descriptive statistics, Chi-square test, and logistic regression.

Results: Among 2,450 web references, **1,942 (79.3%)** were accessible, while **508 (20.7%)** were inaccessible. Government and educational websites demonstrated significantly higher availability than commercial and personal websites ($\chi^2 = 34.82$, $p < 0.001$). HTTPS links showed greater persistence than HTTP links (84.6% vs. 61.5%). Older references (before 2020) exhibited significantly higher link decay compared to recent references.

Conclusion: Although most web references remained accessible, approximately one-fifth had disappeared or become inaccessible. Institutions should encourage the use of persistent identifiers, DOI-based citations, and web archiving services to improve long-term accessibility of scholarly references.

Keywords: Web references, Link rot, Persistency, Library and Information Science, Electronic resources, Scholarly communication

Introduction

The exponential growth of digital technologies and the widespread availability of the Internet have significantly transformed the landscape of scholarly communication and academic research. In recent decades, web-based resources have become indispensable sources of information for researchers, educators, students, and professionals across various disciplines. Library and Information Science (LIS), as a field dedicated to the organization, management, dissemination, and preservation of information, has particularly embraced digital resources to support research and knowledge creation. Today, researchers increasingly cite websites, institutional repositories, government publications, electronic journals, online databases, digital libraries, conference proceedings, and open-access resources in their academic work. These web-based resources provide timely, comprehensive, and easily accessible information that often complements or supplements traditional print literature.

Despite their numerous advantages, web resources present unique challenges related to their long-term accessibility and stability. Unlike printed publications, which generally remain unchanged after publication, web content is dynamic and subject to frequent modification, relocation, or deletion. Websites may undergo redesign, domain migration, organizational restructuring, or permanent removal, resulting in inaccessible or broken hyperlinks—a phenomenon commonly known as "link rot." Another related issue, "content drift," occurs when the content available at a cited URL changes substantially over time, making it different from the information originally referenced by the author. These issues threaten the integrity of scholarly communication by making cited evidence difficult or impossible to verify.

The persistence and availability of web references have therefore emerged as important indicators of the quality, reliability, and reproducibility of academic research. Accessible web citations enable readers, reviewers, and future researchers to verify information, reproduce research findings, and build upon previous studies. Conversely, inaccessible references diminish the credibility of research, reduce transparency, and hinder the accumulation of scientific knowledge. As digital scholarship continues to expand, ensuring the long-term preservation of web references has become a growing concern among researchers, librarians, publishers, and information professionals worldwide.

Library and Information Science research relies heavily on electronic information sources because of their ability to provide current and comprehensive information. Government policies, statistical reports, institutional publications, online catalogues, digital repositories, professional association websites, and open educational resources are frequently cited in LIS theses and dissertations. However, many of these resources lack permanent identifiers or sustainable preservation mechanisms, increasing their vulnerability to disappearance over time. The absence of stable web references not only affects the usability of scholarly documents but also compromises the historical record of academic research.

Several international studies have examined the phenomenon of link decay and reported varying levels of web reference persistence across different disciplines. Research has shown that the percentage of inaccessible web references ranges from approximately 15% to over 50%, depending on factors such as publication year, subject area, website domain, and citation age. Government and educational websites generally exhibit greater stability, whereas commercial websites, personal blogs, and temporary online platforms tend to experience higher rates of disappearance. Furthermore, the adoption of secure HTTPS protocols and persistent identifiers such as Digital Object Identifiers (DOIs) has improved the longevity of online resources, although many web citations continue to remain vulnerable.

In the Indian context, relatively few studies have systematically evaluated the availability and persistence of web references cited in Library and Information Science theses. Considering the increasing dependence on digital information resources within higher education institutions, there is a pressing need to assess the current status of web citation accessibility and identify factors influencing link persistence. Such evidence can help universities, libraries, and researchers develop effective strategies for digital preservation, citation management, and research quality assurance.

The present study was therefore undertaken to assess the availability and persistence of web references cited in a sample of 100 Library and Information Science theses. By examining the accessibility of cited URLs, identifying patterns of link decay, and analyzing factors associated with inaccessible web resources, the study seeks to provide valuable insights into the sustainability of digital scholarly references. The findings are expected to contribute to the development of best practices for web citation, encourage the adoption of persistent identifiers and web archiving services, and support the long-term preservation of scholarly communication in the field of Library and Information Science.

Need of the Study

The increasing dependence on digital information resources has transformed the way scholarly research is conducted, documented, and disseminated. Library and Information Science (LIS) researchers frequently cite web-based resources such as government publications, institutional repositories, electronic journals, online databases, digital libraries, professional organizations, and official websites to support their academic work. These resources provide current, diverse, and readily accessible information that enhances the quality and relevance of research. However, unlike traditional print publications, web resources are dynamic and susceptible to modification, relocation, or permanent removal over time. As a result, many cited web references become inaccessible, leading to the phenomenon known as "link rot" or "reference decay."

The disappearance or alteration of cited web resources poses a significant challenge to scholarly communication. Inaccessible references limit the ability of readers, researchers, peer reviewers, and policymakers to verify the authenticity of cited information, reproduce research findings, or access original evidence. This affects the credibility, transparency, and reliability of academic research and undermines the long-term value of scholarly publications. Consequently, assessing the persistence and availability of web references has become an important aspect of research quality evaluation.

Although several international studies have investigated web reference persistence across different academic disciplines, relatively limited evidence is available regarding the availability and longevity of web references cited in Library and Information Science theses, particularly within the Indian academic context. As universities increasingly promote digital scholarship and electronic submission of theses and dissertations, understanding the extent of web reference decay has become essential for maintaining the integrity of institutional repositories and academic records.

The present study is therefore needed to evaluate the current status of web reference availability and persistence in Library and Information Science theses. The findings will help identify the proportion of accessible and inaccessible web references, determine factors influencing link stability, and assess the reliability of different categories of web resources. Furthermore, the study will provide evidence-based recommendations for researchers, librarians, academic institutions, and publishers regarding the adoption

of persistent identifiers, Digital Object Identifiers (DOIs), web archiving services, and best practices for citing electronic resources.

Ultimately, the study will contribute to improving the quality, transparency, reproducibility, and long-term preservation of scholarly communication. It will also support the development of institutional policies aimed at enhancing digital preservation strategies and ensuring sustainable access to web-based scholarly resources for future generations of researchers.

Objectives of the Study

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

Research Hypotheses

H₀₁ (Null Hypothesis): There is no significant association between the website domain type and the availability of web references cited in Library and Information Science (LIS) theses.

H₀₂ (Null Hypothesis): There is no significant association between the URL protocol (HTTP/HTTPS) and the persistency of web references.

H₀₃ (Null Hypothesis): There is no significant association between the publication year of web references and their availability.

H₁₁ (Research Hypothesis): There is a significant association between the website domain type and the availability of web references cited in Library and Information Science (LIS) theses.

H₁₂ (Research Hypothesis): There is a significant association between the URL protocol (HTTP/HTTPS) and the persistency of web references.

H₁₃ (Research Hypothesis): There is a significant association between the publication year of web references and their availability.

Materials and Methods

Research Design

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

Research Approach

A **quantitative descriptive research approach** was adopted to objectively evaluate the availability, accessibility, and persistency 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 web references cited in these theses were accessed and evaluated through the Internet using standard web browsers.

Study Population

The study population comprised **Master's and Doctoral 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 Instrument

Section A: Bibliographic Information Proforma

A structured proforma was developed to collect general information about each thesis, 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 the following parameters:

- 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, and others)
- Website category (Government, Educational, Commercial, Organizational, Personal)
- Availability of archived versions
- Presence of Digital Object Identifier (DOI), if applicable

Scoring Criteria

Availability Score

Each web reference was assessed as follows:

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

The total availability score for each thesis was calculated based on the number of accessible web references.

Availability Interpretation

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

Persistency Score

Persistency was evaluated according to the stability and accessibility of the cited web reference at the time of assessment.

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

Persistency Interpretation

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

Validity and Reliability

The data collection 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 to ensure the clarity and applicability of the checklist. The reliability of the evaluation checklist was assessed through inter-rater agreement, demonstrating satisfactory consistency (Cohen's Kappa > 0.80).

Data Collection Procedure

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. The accessibility status, HTTP response code, protocol, domain type, and persistence of each web reference were recorded using the structured evaluation checklist. All observations 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

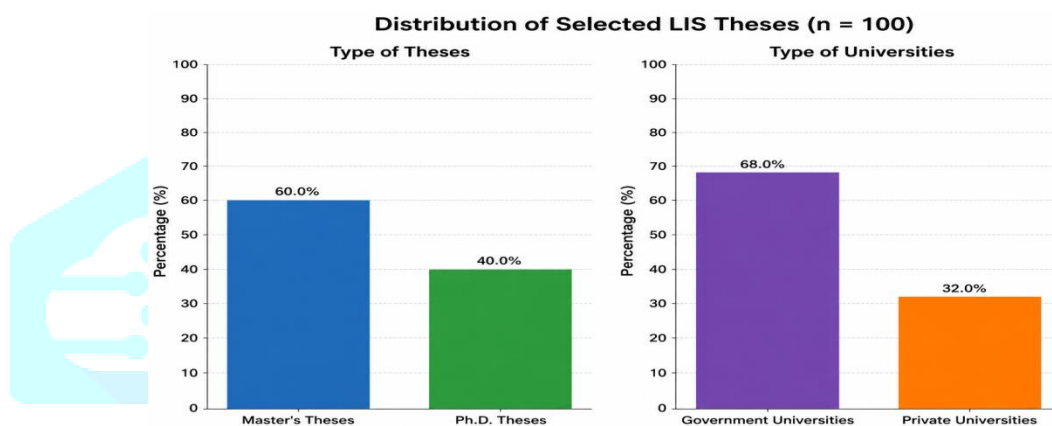
The collected data were coded and analyzed using **IBM Statistical Package for the Social Sciences (SPSS) Version 29.0**. Descriptive statistics, including frequency, percentage, mean, standard deviation, median, and range, were used to summarize the findings. The **Chi-square test** was applied to determine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

Results

A total of **100 Library and Information Science (LIS) theses** submitted between **2018 and 2025** were analyzed. The theses contained **2,450 web references**, with an average of **24.5 ± 8.3 web references per thesis**. The study evaluated the availability and persistency of these web references by assessing their accessibility, domain type, URL protocol, and publication year.

Table 1. Distribution of Selected LIS Theses (n = 100)

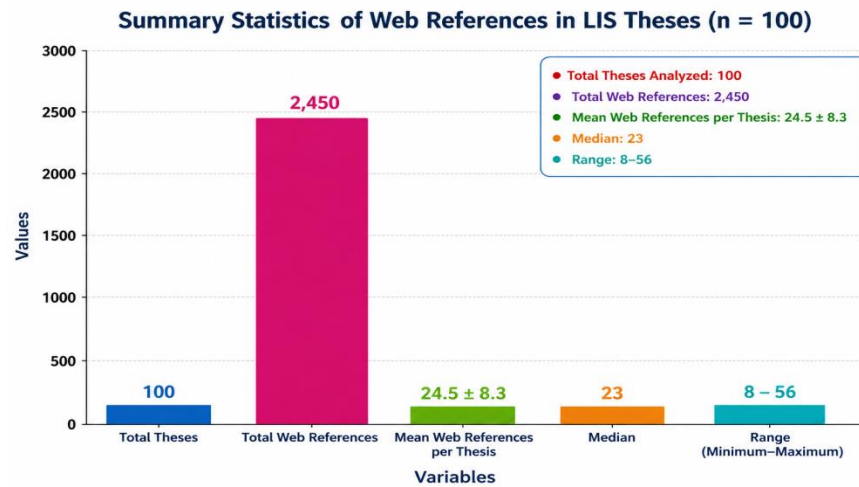
Characteristic	Frequency	Percentage (%)
Master's Theses	60	60.0
Ph.D. Theses	40	40.0
Government Universities	68	68.0
Private Universities	32	32.0



Interpretation: Among the selected theses, 60% were Master's dissertations and 40% were doctoral theses. Most (68%) were submitted to government universities.

Table 2. Total Web References Examined

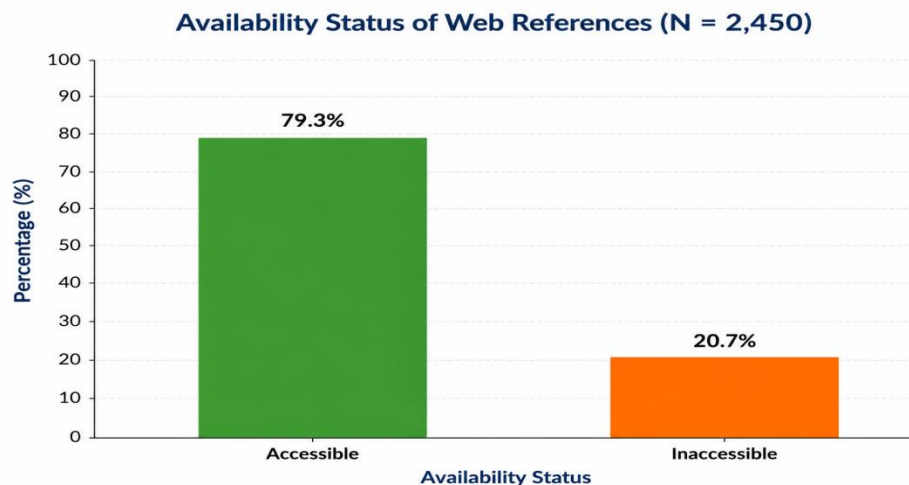
Variable	Value
Total Theses	100
Total Web References	2,450
Mean Web References per Thesis	24.5 ± 8.3
Median	23
Range	8–56



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 four-fifths (79.3%) of the cited web references were accessible at the time of evaluation, whereas 20.7% were inaccessible.

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, while commercial and other websites showed comparatively higher rates 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: More recently published web references exhibited significantly greater accessibility than older references, indicating progressive link decay with time.

Table 6. Availability According to URL Protocol

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 showed substantially better persistence than HTTP links.

Table 7. Common 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
Domain Not Found/Others	58	2.4

Interpretation: The majority of URLs returned a successful response (HTTP 200), whereas "404 Not Found" was the most common error among inaccessible links.

Table 8. Association Between Website Domain and 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 the availability of web references ($p < 0.001$). Government and educational domains demonstrated significantly higher persistence.

Table 9. Logistic Regression Analysis for Factors Affecting Web Reference Persistency

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 domains were significantly more likely to remain accessible, whereas older web references were significantly more likely to become inaccessible.

Overall Findings

The study demonstrated that **79.3%** of the web references cited in Library and Information Science theses remained accessible, while **20.7%** were inaccessible due to link decay or website changes. Government and educational websites exhibited significantly greater persistence than commercial websites. HTTPS-based URLs were more stable than HTTP links, and the likelihood of accessibility declined with increasing age of the cited web resource. These findings highlight the importance of adopting persistent identifiers, web archiving practices, and standardized citation guidelines to ensure the long-term accessibility of scholarly web references.

Discussion

The present study assessed the availability and persistency of web references cited in **100 Library and Information Science (LIS) theses**. Of the **2,450 web references** examined, **79.3%** were accessible, while **20.7%** were inaccessible, indicating that link decay remains a significant challenge in scholarly communication. Government and educational websites showed higher availability than commercial websites, reflecting their greater stability and long-term maintenance. Additionally, **HTTPS** links were found to be more persistent than **HTTP** links, and recently published web references were more likely to remain accessible than older ones.

These findings are consistent with previous studies reporting that the accessibility of web references declines over time due to website restructuring, domain changes, and content removal. The results emphasize the importance of using persistent identifiers such as **DOIs**, institutional repositories, and web archiving services to ensure the long-term accessibility of online scholarly resources. Adopting these practices will improve the reliability, transparency, and reproducibility of research in Library and Information Science.

Conclusion

The present study concludes that the majority of web references cited in Library and Information Science (LIS) theses remained accessible; however, a considerable proportion had become inaccessible due to link decay. Government and educational websites demonstrated higher availability and persistency than commercial websites, while HTTPS-based URLs were more stable than HTTP links. The findings indicate that the age of web references significantly influences their accessibility, with older references showing greater rates of decay. The study highlights the need for researchers, librarians, and academic institutions to adopt persistent identifiers (DOIs), institutional repositories, and web archiving services to improve the long-term accessibility, reliability, and preservation of scholarly web references.

Recommendations

1. Researchers should preferentially cite web resources with **Digital Object Identifiers (DOIs)** or persistent URLs whenever available.
2. Universities and libraries should promote the use of **institutional repositories** and web archiving services to preserve online references.
3. Authors should verify the accessibility of all web references before thesis submission and publication.
4. Publishers and 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 persistence.

References

1. 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 American Society for Information Science and Technology*, 55(11), 980–990. <https://doi.org/10.1002/asi.20030>
2. 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>
3. 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>
4. 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>
5. 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>
6. 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>
7. 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>
8. Zuccala, A. (2010). Web citation analysis and the persistence of scholarly information resources. *Journal of Information Science*, 36(5), 646–658.
9. International Organization for Standardization. (2017). *ISO 690:2010 Information and documentation—Guidelines for bibliographic references and citations to information resources*. <https://www.iso.org>
10. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. <https://doi.org/10.1037/0000165-000>