

Reference rot degrades information preservation and induces the loss of intellectual integrity

■ JAIME A. TEIXEIRA DA SILVA¹

Abstract

The age of open access has ushered in a greater desire to cross-cite information from a multitude of sources, some of which may have a determined fate and life cycle. Information insecurity caused by the loss or transposition of information also negatively impacts information integrity by reducing its use and usefulness. Reference rot refers to the phenomenon in which the link to a web resource or journal article URL no longer functions, revealing instead a “404 not found” error message. Reference rot can reduce the reliability and usefulness of a manuscript because access to information supporting claims and/or positions within a paper cease to exist. Academic papers carry a complex mixture of information that is derived from a multitude of sources. Collectively, they ensure a paper’s health and functionality, aspects that fade as access to supporting information becomes truncated, i.e., reference rot, ultimately reducing the usefulness of the academic paper, and making it, and its claims, unreliable. Although it is a cumbersome task, as the curators of academic and scientific information, extant journals and their editors should revisit URLs in the reference lists regularly to update any broken links or URLs, and correct reference lists accordingly. This laborious task should involve close coordination between editors and authors to ensure, as best as possible, the sustained integrity of citations and thus the information backbone of a manuscript. An academic paper with a strong, or fortified, citation base, has greater information integrity, reliability and use for science and society.

Keywords: citations; information; literature; metrics; reliability; trust

Dear *Epistēmēs Metron Logos* Editors,

References and citations do not exist in a vacuum: information links

The world of open access (OA) has opened up academia to a world of opportunities, but has also left it vulnerable to a parallel world of threats (Al-Khatib and Teixeira da Silva, 2017). These threats have exposed academics, their institu-

tions and their funders to an unimaginable range of predatory and exploitative tactics (Teixeira da Silva et al., 2019). When academics are under threat, then so too are academia and academic societies (Hewitt et al., 2017). While open knowledge has certainly empowered the sharing and creation of new knowledge, leading to its greater democratization (Holbrook, 2019), like many societal ills, it has also attracted the attention of abusive and disruptive elements that threaten its integrity. One way in which the exploitation of the “open” brand is taking place in the gold OA model is by unabated and unregulated mining of wealth under the pretext of the “goodness” that comes from all academics being able to see and access

¹Independent researcher, P. O. Box 7, Miki-cho post office, Ikenobe 3011-2, Kagawa-ken, 761-0799, Japan. jaimetex@yahoo.com

information “freely” and openly, but with a catch, article publishing charges (APCs).

Despite this tectonic shift to an increasingly profitable APC-based business model (Björk, 2017), and the deafening clamors by the same OA proponents to make OA an integrity-based open science model supported by artificial intelligence (AI) (Burgelman et al., 2019), these tactics may simply be a well-greased marketing-based ruse to extract more wealth, for there is much of it to be extracted from academia. Somewhere along that road to wealth and “openness”, in their quest for “fool’s gold”, OA policy-makers and thought-leaders curiously seem to have – perhaps conveniently – overlooked basic fundamental aspects of transparency regarding APCs (Teixeira da Silva, 2020a). Salvation from this self-imprisoning and exploitative information ecosystem could be as diamond or platinum OA (Hiemstra et al., 2020), which incurs no costs to academic institutes or taxes on authors for their intellectual contribution, but which, by virtue of real operation costs, is not a sustainable publishing model, and there is always the risk of imminent collapse. Another risk of APC hyperinflation (Khoo, 2019) at the expense of intellectual deflation is the loss of information, which can be as spontaneous as it can be induced.

While most of these pro-open movements advocate, almost in a state of blind mesmerization, growth, expansion, fortification and diversification, almost oblivious of the true weighting of these costs, a blind spot is getting increasingly larger: the loss of information. OA proponents will surely argue that greater access to information is of benefit to society, but insufficient attention is being paid to the loss of information, and the concurrent impact on integrity caused by that loss. Is all information useful, is all information valid, and does all information have integrity? While this topic in itself would likely fill the pages of volumes of books, turning to humanity’s current predicament, COVID-19, provides a perfect example to appreciate the melting pot of degrading values in information safety (i.e., misinformation) and integrity (i.e., fake or pseudo-science) amidst a clamor for more stringent information integrity (Teixeira da Silva, 2020b).

At this crossroad between information integrity and preservation, a key question needs to be asked: should the loss of useful information be given equal status and emotive attention as the loss of useless information? If information could be simply characterized into a bimodal state, as true or false or a binary 0 or 1, then it would be so easy to separate the wheat (useful information) from the chaff (useless information). Yet, in the OA cloud, wheat and chaff are mixed, and truth and false exist in many overlapping shades, making their distinction imprecise, if not impossible (Sanbonmatsu and Johnston, 2019). It would be easy to emotively argue for – on one extreme – the loss of information that is false (misinformation), deceitful, fraudulent, insulting or opaque, or – on the other extreme – for the preservation of “sound” information. Yet, the existence of imprecisions – perhaps hidden or undiscovered – in the latter body of intellect could reverse its characterization, placing it into the same category as the former set of information. Thereby, by a mere transition of intellectual status, information’s security is endangered and its integrity is threatened. Even so, while information security and integrity are collated, they have distinct phenomena with equally distinct fortification measures (Clarkson and Schneider, 2015; Chatzikokolakis, 2018). While few might advocate for the preservation of “bad science”, even fewer should advocate for the irretrievable loss of information.

In academic publishing, the intersection between threats to information security and validity, and a non-ethics-based appreciation of integrity, appears to have received little attention, apart from the field of “predatory publishing” (Teixeira da Silva et al., 2019). However, the loss of information due to the failure of AI, in turn caused or spurred by human fallibility, might trigger a new cycle of technology panic (Orben, 2020) and a veritable crisis may arise, not unlike the current crisis in trust caused by theoretical lapses, leaving academia in desperate need of an epistemological revolution (van Rooij and Baggio, 2020).

Within this prism of evolving and devolving information where integrity and security are fluid, in part the result of a mix of intellectual strife, opportunism, carelessness and detachment, even as

both useful and useless information get rewarded, that academics find themselves in a practical bind. It is very frustrating for scholars to turn to a source of information, either as a reference or website link, only to find that it no longer exists, or to realize that the link is dead, and to be met with a “404 not found” error message. Sometimes there are valid reasons for spasmodic access to defunct links, including websites that have migrated, whose servers have been modified, delisting from indexed sites, the discontinuation of a journal and its host website (Laakso et al., 2020), or more nefarious reasons, such as the need to change or shut down a website in response to a journal hijacking (Trapp, 2020). Except for erroneously typed uniform resource locator (URL) addresses that will undoubtedly reveal a 404 error, the vast majority of such dead links are unexpected events that even authors who originally cite them cannot envision, or prevent. In some ways, these dead links reveal a wider failure in the dynamics of human-created technological development (Lorenz, 2017), namely the inability of AI to detect and automatically correct errors that were created by humans but supported by AI-supported publishing-based technologies. For this reason, academics need to select very carefully the sources of their information (Nicholas et al., 2015), keeping in mind that URL-based links and sources could change, or vanish, at any instant.

Citation practices not only speak to the great care that academics need to have when selecting sources of information, especially those that are exclusively online, but to a wider ill in the culture of academic publishing, namely the phenomenon of the pressure felt by academics to publish quickly and in greater volumes, either to meet peer pressure or to satisfy work-related expectations (Frith, 2020). The latter “pressure” may lead academics to adopt lax rigor in their selection of literature as a result of the unavailability of sufficient support, help infrastructure, or advice (Giblin et al., 2021), and insert sources of information that might not be reliable simply because they are under pressure to offer a citation for any and all claims in a paper, or to select legitimate sources that, through no fault of their own, simply vanish due to technological failure.

Reference rot: the intersection between citation invalidation and information degradation

The term “reference rot” (RR), which can be applied to both journal references and web resources, was coined to represent two events: link rot, in which the content of a reference can no longer be accessed since the URL, or uniform resource identifier (URI) is no longer functional, or content drift, in which the content at a specified URI changes over time and no longer represents the original content (Klein et al., 2014). RR threatens the integrity of the scholarly record and is not merely a nuisance or irritant to academics because it disrupts information flow and access to knowledge. This threat caused by RR is expected to get larger as OA expands and online content grows, stressing the need for the long-term digital preservation of scholarly content (Kirchhoff, 2008; Lynch, 2017).

Select attempts have been made to quantify RR. Klein et al. (2014) examined RR frequency in a 15-year period (1997-2012) in more than 3.5 million articles derived from *arXiv*, Elsevier, and PubMed Central, 1.8 million of which referenced in excess of 1 million web resources. They found that the incidence of RR from these three sources was 13%, 22%, and 14% in 2012, 18%, 41%, and 36% in 2005, and 34%, 66%, and 80% in 1997, respectively, showing a massive increase over time. There was 34.1% RR in *Emergency Medicine Australasia* papers (2010-2014) (O’Connor and O’Connor, 2015), 34.0% RR in *Irish Medical Journal* papers (2013-2017) (O’Connor and O’Connor, 2018), and 27.3% RR after 13-25 months in URLs of 2013 *International Urogynecology Journal* papers (Riss, 2015). These findings suggest that information in these references and journals becomes increasingly unreliable over time due to RR, so access to information originally held in links from reference lists can no longer be reliably accessed. This in turn reduces the integrity of papers that suffer from RR, and limits their usefulness and ability to be cited.

In the context of information erosion, it is understandable that, over time, that some – maybe even many – URLs and URIs in academic papers, including in their reference lists and footnotes, may change location, or disappear altogether. Although this is not a common academic require-

ment of editors (Teixeira da Silva and Dobránszki, 2018), as journal integrity gate-keepers, it is incumbent upon editors and journals/publishers to regularly (for example, annually) revisit their published literature in extant journals to ensure that RR and content drift has not occurred, perhaps by assigning a digital object identifier (DOI) to each URL/URI and to archive papers upon submission (Burnhill et al., 2015). In the face of RR, editors should issue an erratum that updates, or annuls, cases of RR so as to fortify the integrity of the published literature, although such a correction of the literature may become very cumbersome if RR occurs periodically, so suitable solutions that address this issue are needed. One way to achieve the posterity of information is through the archival of a website on platforms such as the Internet Archive (<https://archive.org/index.php>) or CLOCKSS (<https://clockss.org/>). There is also the need to develop digital libraries that are able to trace the flow of information, as part of its integrity management (Hedin et al., 2017).

Academic citation practices, like many fields of study in academia, are in a dynamic state of change (Hyland and Jiang, 2019), and the ease with which information can be digitally cross-linked to other sources of information, either through simple web searches such as on Google or Google Scholar, or refined searches on dedicated information-centralized platforms and databases like PubMed, Web of Science or Scopus, has fortified the growth of academic knowledge and empowered academics to seek and share knowledge in unlimited and unfiltered ways, even more so in an OA playing field. Valid information in a link today can, by human fallibility and technological (AI) failure, be invalid tomorrow. Unlike ethics-based threats to the integrity of the literature that rely on deceitful practices to delegitimize academic publishing, such as paper mills (Teixeira da Silva, 2021) or the creation of fake identities (or their creators) with the purposeful intent of abusing scholarly trust and denigrating academic publishing value systems (Teixeira da Silva, 2020c), RR is a unique technology-based degradation of academic publishing's integrity, but that relies on human failure to correct it. While the former cases of ethics-based threats to academia's integrity are voluntarily induced, the latter case

(RR) is involuntary. In either case, robust solutions to digital preservation (Lynch, 2017) can only arise when human intelligence and AI combine to resolve these issues while removing the threats, or offering solutions to protect academia while ensuring or fortifying the integrity of the literature.

Acknowledgement

The author thanks the advice offered by Colm O'Connor (RCSI Library, Dublin, Ireland) on an earlier version of the paper.

Conflicts of interest

The author declares no conflicts of interest of relevance to this topic.

Author contributions

The author contributed entirely to the intellectual discussion underlying this paper, literature exploration, writing, reviews and editing, and accepts responsibility for the content of the paper.

References

- Al-Khatib, A., Teixeira da Silva, J.A. (2017). Threats to the survival of the author-pays-journal to publish model. *Publishing Research Quarterly* 33(1): 64–70. <http://doi.org/10.1007/s12109-016-9486-z>
- Björk, BC. (2017). Scholarly journal publishing in transition- from restricted to open access. *Electronic Markets* 27(2): 101–109. <https://doi.org/10.1007/s12525-017-0249-2>
- Burgelman, J-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., Schoupe, M. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data* 2: 43. <https://doi.org/10.3389/fdata.2019.00043>
- Burnhill, P., Mewissen, M., Wincewicz, R. (2015). Reference rot in scholarly statement: threat and remedy. *Insights* 28(2): 55–61. <http://doi.org/10.1629/uksg.237>
- Chatzikokolakis, K. (2018). On the additive capacity problem for quantitative information flow. In: McIver A., Horvath A. (eds) *Quantitative Evaluation of Systems. QEST 2018. Lecture Notes in Computer Science*, vol

11024. Springer, Cham, Switzerland, pp. 1–19. https://doi.org/10.1007/978-3-319-99154-2_1
- Clarkson, M., Schneider, F. (2015). Quantification of integrity. *Mathematical Structures in Computer Science* 25(2): 207–258. <http://doi.org/10.1017/S0960129513000595>
 - Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences* 24(1): 1–2. <http://doi.org/10.1016/j.tics.2019.10.007>
 - Giblin, J., Stefaniak, J., Eckhoff, A., Luo, T. (2021). An exploration of factors influencing the decision-making process and selection of academic help sources. *Journal of Computing in Higher Education* 33(1): 1–18. <https://doi.org/10.1007/s12528-020-09252-0>
 - Hedin D., Sjösten A., Piessens F., Sabelfeld A. (2017). A principled approach to tracking information flow in the presence of libraries. In: Maffei M., Ryan M. (eds) *Principles of Security and Trust*. POST 2017. Lecture Notes in Computer Science, vol 10204. Springer, Berlin, Heidelberg, pp. 49–70. https://doi.org/10.1007/978-3-662-54455-6_3
 - Hewitt, M., Dingwall, R., Turkmendag, I. (2017). Squeeze on academics poses threats to learned societies. *Research Fortnight* 28 June 2017: 23–23. <https://eprint.ncl.ac.uk/239545>
 - Hiemstra, D., Moens, M.F., Perego, R., Sebastiani, F. (2020). Transitioning the information retrieval literature to a fully open access model. *ACM SIGIR Forum* 54(1): 1–10. <http://sigir.org/wp-content/uploads/2020/06/p15.pdf>
 - Holbrook, J.B. (2019). Open science, open access, and the democratization of knowledge. *Issues in Science and Technology* 35(3): 26–28. <https://issues.org/wp-content/uploads/2019/06/Holbrook-Philosophers-Corner-SP19-ISSUES-26-28.pdf>
 - Hyland, K., Jiang, F.K. (2019). Points of reference: changing patterns of academic citation. *Applied Linguistics* 40(1): 64–85. <https://doi.org/10.1093/applin/amx012>
 - Khoo, S.Y.-S. (2019). Article processing charge hyperinflation and price insensitivity: an open access sequel to the serials crisis. *LIBER Quarterly* 29(1): 1–18. <http://doi.org/10.18352/lq.10280>
 - Kirchhoff, A.J. (2008). Digital preservation: challenges and implementation. *Learned Publishing* 21(4): 285–294. <https://doi.org/10.1087/095315108X356716>
 - Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., Tobin, R. (2014). Scholarly content not found: one in five articles suffers from reference rot. *PLoS ONE* 9(12): e115253. <http://doi.org/10.1371/journal.pone.0115253>
 - Laakso, M., Mathias, L., Jahn, N. (2020). Open is not forever: a study of vanished open access journals. *Arxiv* (preprint, not peer reviewed, version 3): 1–32. <https://arxiv.org/abs/2008.11933>
 - Lorenz, S. (2017). Ecological criticism of growth and the means and ends of technology. A pragmatist perspective on societal dynamics. *Journal of Cleaner Production* 166: 98–106. <https://doi.org/10.1016/j.jclepro.2017.08.008>
 - Lynch, C. (2017). Updating the agenda for academic libraries and scholarly communications. *College & Research Libraries* 78(2): 126–130. <https://doi.org/10.5860/crl.78.2.126>
 - Nicholas, D., Jamali, H.R., Watkinson, A., Herman, E., Tenopir, C., Volentine, R., Allard, S., Levine, K. (2015). Do younger researchers assess trustworthiness differently when deciding what to read and cite and where to publish? *International Journal of Knowledge Content Development & Technology* 5(2): 45–63. <https://doi.org/10.5865/IJKCT.2015.5.2.045>
 - O'Connor, C., O'Connor, A.E. (2015). 'Reference rot': A developing problem in *Emergency Medicine Australasia*. *Emergency Medicine Australasia* 27(5): 494. <https://doi.org/10.1111/1742-6723.12459>
 - O'Connor, C., O'Connor, A.E. (2018). Reference rot in medical publications. *Irish Medical Journal* 111(9): 827. PMID 30556675
 - Orben, A. (2020). The Sisyphean cycle of technology panics. *Perspectives on Psychological Science* 15(5): 1143–1157. <https://doi.org/10.1177/1745691620919372>
 - Riss P. (2015). "Reference rot": does it matter? *International Urogynecology Journal* 26(9): 1251–1252. <http://doi.org/10.1007/s00192-015-2707-8>
 - Sanbonmatsu, D.M., Johnston, W.A. (2019). Redefining science: the impact of complexity on theory development in social and behavioral research. *Perspectives on Psychological Science* 14(4): 672–690. <https://doi.org/10.1177/1745691619848688>
 - Teixeira da Silva, J.A. (2020a). Three new suggested guidelines for increased transparency regarding open access article processing charges (APCs). *Epistēmēs Metron Logos* 4: 4–7. <https://doi.org/10.12681/eml.24208>
 - Teixeira da Silva, J.A. (2020b). Misinformation in COVID-19 media and literature, with an emphasis on

open data policies. *Journal of Advocacy, Research and Education* 7(2): 25–29. http://www.kadint.net/journals_n/1611170173.pdf

• Teixeira da Silva, J.A. (2020c). Is there a need for creators of imaginary authors to face legal consequences? *Croatian Medical Journal* 61(6): 561–563. <https://doi.org/10.3325/cmj.2020.61.561>

• Teixeira da Silva, J.A. (2021). Paper mills and on-demand publishing: Risks to the integrity of journal indexing and metrics. *Medical Journal Armed Forces India* 77(1): 119–120. <https://doi.org/10.1016/j.mjafi.2020.08.003>

• Teixeira da Silva, J.A., Dobránszki, J. (2018). Editors moving forward: stick to academic basics, maximize transparency and respect, and enforce the rules. *Recenti Progressi in Medicina* 109(5): 263–266. <https://doi.org/10.1701/2902.29244>

• Teixeira da Silva, J.A., Dobránszki, J., Tsigaris, P., Al-Khatib, A. (2019). Predatory and exploitative behaviour in academic publishing: An assessment. *The Journal of Academic Librarianship* 45(6): 102071. <https://doi.org/10.1016/j.acalib.2019.102071>

• Trapp, J. (2020). Predatory publishing, hijacking of legitimate journals and impersonation of researchers via special issue announcements: a warning for editors and authors about a new scam. *Physical and Engineering Sciences in Medicine* 43(1): 9–10. <https://doi.org/10.1007/s13246-019-00835-5>

• van Rooij, I. Baggio, G. (2020). Theory development requires an epistemological sea change. *Psychological Inquiry* 31(4): 321–325. <https://doi.org/10.1080/1047840X.2020.1853477>