Developing an Archiving Strategy: Archiving outbound links?

As part of the OBF project the Garth Stewart at the Digital Preservation Coalition undertook a report looking specifically at issues around the archiving and preservation of outgoing links and the content associated with those links.

Stewart, G. (2026). Think, before you link: Link Rot, eBooks, and digital preservation. Zenodo. https://doi.org/10.5281/zenodo.19912435

The report provides helpful information for both publishers and authors to help them realistically mitage against the reprecussions of link rot for their publications. Key recommendations for authors and publishers, as laid out in Steps 1 to 5 of Part B of the Report, are:

1. Think before you link, and determine acceptable loss

Link rot is inevitable, so authors and publishers must be selective about which links they include. Not all links are equally valuable—some are critical to understanding the work, while others are supplementary.

Stakeholders should assess whether a link “must survive” and define an acceptable level of loss. This involves prioritising essential content, avoiding unnecessary links, and considering the longevity and reliability of the source. Authors should also evaluate how broken links would affect user experience, trust, and the integrity of the work.

Where possible, critical external content should be embedded or packaged with the eBook. Ultimately, this step promotes a risk-based, intentional approach to linking, recognising that some loss is normal and manageable if planned for.

2. Cite to resilient links

Focuses on improving the durability of links that are included. Authors should use simple, stable, and transparent URLs, avoiding long, complex, or shortened links that are prone to failure. Preference should be given to trusted sources, landing pages rather than deep file paths, archived and open-access content.

Persistent identifiers (e.g. DOIs) are encouraged where available, though these should not be assumed to be permanent.

Citing archived versions of webpages (e.g. from web archives) is best practice for critical content, as these are more stable and preserve content in a fixed state. Clear and consistent citation also acts as “preservation metadata,” helping future users locate content even if links break.

3. Describe more, bolster accessibility

To mitigate the impact of link rot, authors should provide detailed descriptions of linked content, including its context, purpose, and structure. This can include summaries, transcripts, or screenshots, especially for multimedia resources. Even brief descriptions can preserve meaning if the link fails. This approach not only strengthens long-term resilience but also improves accessibility for users and machines. Documentation can be embedded in the text or stored as metadata.

While this requires additional effort, it is far less costly than attempting to recover lost content later.

Overall, richer description ensures that the intellectual value of the work remains understandable even when links disappear.

4. Archive your URLs

For important or critical links, stakeholders should proactively create archived copies using web archiving tools or services.

These snapshots (e.g. via services like “Save Page Now” or similar tools) provide stable, time-stamped versions of content that can be cited or stored alongside the eBook. Where possible, multiple copies or formats (e.g. WACZ/WARC files) should be created to increase redundancy. These files could also be archived alongside the content and metadata files in the folder used for archiving the publication itself.

These archived versions can either replace or supplement live links, reducing dependence on the original source. Although this step requires more effort and resources, it significantly increases long-term access and protects against both link rot and content drift.

5. Monitor your URLs, enact digital archaeology

After publication, links should be periodically checked to identify breakage. Publishers and repositories can use automated tools to monitor link health, particularly for online eBooks. For open access works it is also not unusual for readers to report broken links when these are observed. When links fail, “digital archaeology” involves searching web archives or other sources to recover or replace missing content. While not all links can be restored, this process can often recover valuable material. This step acknowledges that preservation is ongoing, not one-time, and requires maintenance over time. It is especially important for high-value links, helping sustain the usability, credibility, and completeness of the eBook as it ages.

Revision #4
Created 2026-04-28 15:12:11 UTC by Rupert Gatti
Updated 2026-04-30 17:32:20 UTC by Rupert Gatti