



You think the reference in that online article links to what the author referenced? Think again!

January 5, 2017

It happens all the time. You are looking at a scientific article you found and there is a link embedded in the text that references an important part of your needs. You click on it and get the dreaded error message – 404 not found - that the page required cannot be accessed. At least this error message is unambiguous: the content is unavailable. But what if you smoothly end up at the linked web page? Is the content of that page still the same as it was when the link was made?

As publishing becomes more web-based, understanding the impact that the web's dynamic and ephemeral nature has on the scientific record becomes more important, and scientists from the Los Alamos National Laboratory are leading the way. According to a paper they published in Public Library of Science One (PLOS ONE) on December 2, 2016, "Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content", web dynamics yield two different phenomena that have a detrimental impact. Link Rot refers to the disappearance of the referenced web page altogether. The second phenomenon is Content Drift, where a reference points to a web page that has changed since the original paper was published and hence no longer represents the content that existed when the author referenced it. These two phenomena combined are referred to as Reference Rot.

Prior studies, including one conducted in 2014 by the same team from Los Alamos, largely focused on quantifying Link Rot. This new paper is the first to scientifically quantify Content Drift for references to web pages made in scientific articles. It does so by first selecting representative snapshots of referenced pages from web archives around the world, and then textually comparing these snapshots with their counterparts on the live web.

"We were astonished to find that, for those links that still worked, more than 75 percent led to content that was different from what was originally referenced," said Herbert Van de Sompel, of the Los Alamos National Laboratory Research Library. "This is especially disconcerting because readers who follow these links are totally unaware that they are not retrieving the content that the author referenced."

Los Alamos authors Shawn M. Jones, Herbert Van de Sompel, Harihar Shankar, and Martin Klein, all working at the Laboratory's Research Library, revisited the dataset from their 2014 PLOS ONE paper to quantify the extent of Content Drift. They found that, overall, 75% of references to web pages are affected and that the problem gets

progressively worse the older the referencing article is. They also found that archived snapshots that are representative of what the author referenced are only available for 30% of references. These numbers are alarming and provide a unique insight in the impact that web dynamics have on the integrity of the web-based scholarly record. Fortunately, the authors also studied approaches to ameliorate the Reference Rot problem, and, with this regard, advocate the use of [Robust Links](#).

The article was coauthored by Richard Tobin and Claire Grover from the University of Edinburgh, a partner in the Hiberlink project, an international effort funded by the Andrew W. Mellon Foundation focused on addressing Reference Rot.

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Managed by Triad National Security, LLC for the U.S Department of Energy's NNSA

