

Unavailability of online supplementary scientific information from articles published in major journals

Evangelos Evangelou^{*,1} Thomas A. Trikalinos,^{*,†} and John P.A. Ioannidis,^{*,†,‡}

^{*}Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; [†]Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Boston, Massachusetts, USA; and [‡]Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece

ABSTRACT Printed articles increasingly rely on online supplements to store critical scientific information, but such data may eventually become unavailable. We checked the current availability of online supplementary scientific information published in six top-cited scientific journals (*Science*, *Nature*, *Cell*, *New England Journal of Medicine*, *Lancet*, *Proceedings of the National Academy of Sciences USA*). Here we show that in 4.7% and 9.6% of articles with online supplementary material, some of the supplements became unavailable within 2 and 5 years of their publication, respectively.—Evangelou, E., Trikalinos, T. A., Ioannidis, J. P. A. Unavailability of online supplementary scientific information from articles published in major journals. *FASEB J.* 19, 1943–1944 (2005)

Key Words: supplementary information • broken links • internet • scientific journals

Given page and cost limitations, much critical information on published scientific articles appears increasingly only as online supplementary information. This material includes key methodological details, data, tables, images, and video clips and may be a sine qua non for the critical appraisal, documentation, replication, and further exploitation of published research. Print versions of articles are becoming increasingly shorter and more elliptical given the convenience of web posting (1). However, might online supplements eventually be lost and become unavailable to the interested scientists? In previous studies, “broken links” have already been found to be a problem for online references, educational material, and web pages in general (2–5).

We checked the current availability of online supplementary scientific information published 2 and 5 years ago in the six top scientific journals that receive > 120,000 citations annually and have journal impact factors > 10 according to the 2003 edition of the Journal Citation Reports issued in the Thompson-Institute for Scientific Information Web of Knowledge (*Science*, *Nature*, *Cell*, *New England Journal of Medicine*,

Lancet, *Proceedings of the National Academy of Sciences USA*) (6). We probed all articles published in March 2000 and 2003 that included links pointing to any type of supplementary online information, except for general reference links.

We focused our screening on papers that represented original research. This included research articles, brevia, reports, and technical comments in *Science*; articles, letters, and brief communications in *Nature*; all regular articles in the *Proceedings of the National Academy of Sciences*; original articles and special articles in the *New England Journal of Medicine*; articles, research letters, and mechanisms of disease articles in *Lancet*; and all regular articles in *Cell*.

We recorded the following information from each eligible article: authors, title, scientific domain, journal of publication, and number of active and broken web links directing to supplementary material. We also recorded, judging from the link’s address, whether the supplementary information was stored in the pertinent journal’s web servers (servers belonging to the journal or the publisher) or in other web pages (personal web pages or web spaces belonging to the authors’ affiliated institutions and organizations).

We did not consider general reference links, such as links referring to National Center for Biotechnology Information databases and sequence databases in the public domains.

Link verification was done with the specialized free-ware software *Xenu’s Link Sleuth* (ver 1.2, Tilman Hausherr, 2005), which can check links and report their status. Links pointing to supplementary scientific material were checked in March and April 2005. Information was considered missing if links could not be accessed in two consecutive attempts 2 weeks apart during this period. Links were checked at varied times of the day and at different days of the week to avoid

¹ Correspondence: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece. E-mail: me01760@cc.uoi.gr
doi: 10.1096/fj.05-4784sf

TABLE 1. Unavailable online supplementary information

Journal (year)	Broken/total supplements		Specific articles (volume; pages [field]) ^a
	Articles	Links	
Nature (2000)	2/15	5/23	404;385–387 (Conservation Biology) 404;398–402 (Molecular Genetics)
Science (2000)	0/30	0/33	None
Cell (2000)	0/1	0/4	None
NEJM (2000)	0/0	0/0	None—no articles published with links
Lancet (2000)	1/3	1/5	355;1064–1069 (Medicine/Clinical trials)
PNAS (2000)	4/24	6/46	97;2450–2455 (Applied Biological Sciences) 97;2562–2566 (Biophysics) 97;2680–2685 (Genetics) 97;3509–3514 (Microbiology)
Nature (2003)	1/30	1/80	422;317–322 (Cell Biology)
Science (2003)	0/41	0/55	None
Cell (2003)	0/5	0/30	None
NEJM (2003)	2/5	2/10	348;977–985 (Medicine/Epidemiology) 348;1223–1232 (Medicine)
Lancet (2003)	2/15	3/49	361;918–922 (Medicine/Ethics) 361;923–929 (Medicine/Microarrays)
PNAS (2003)	3/75	3/250	100;2807–2812 (Microbiology) 100;3233–3238 (Cell Biology) 100;3293–3298 (Developmental Biology)

^a Some (not necessarily all) of the article links with online supplementary information were broken.

routine server maintenance. All inactive links were manually verified as well.

Of 955 screened articles, 244 articles with links ($n=585$ links) qualified. There was considerable variability in the extent of use of online links to post supplementary information across these six journals (Table 1). Online supplements were particularly prevalent in *Science*, *Nature*, and the *Proceedings of the National Academy of Science* but had become more common in 2003 compared with 2000 for the other three journals as well.

Online supplementary information could not be retrieved for 21 links in 15 articles (6.2%) (Table 1). Seven of 73 papers (9.6%) contained 12 broken links in 2000 and 8 of 171 articles (4.7%) contained 9 broken links in 2003. The entire online information was missing in 8.2% and 1.8% of the articles with links published in 2000 and 2003, respectively.

Broken links were encountered in four of the six journals, and their articles pertained to a wide variety of scientific topics (Table 1). The proportion of articles with broken links was highest in the two high-profile medical journals (33% [$1/3$] in 2000 and 20% [$4/20$] in 2003). Of the 21 broken links, 14 were referring to web pages not belonging to the journal or publisher, and these apparently were personal or institutional web pages; seven broken links were hosted in journal/publisher servers.

These results indicate that even in the most prestigious and visible journals, some scientific information may eventually become unavailable when it is supplied only online; personal and institutional web pages may be particularly vulnerable. Here we examined quite

recent publications, and the proportion of lost information may increase as time passes unless proper measures are taken to remedy the situation. It should not be taken for granted that investigators would be able to furnish again this unavailable information (3, 7). Given that supplementary information is an essential part of a published article, journals and publishing groups need to improve their methods to ensure the maintenance and continuous availability of this important scientific material. FJ

REFERENCES

1. Delamothe, T., Mullner, M., and Smith, R. (1999) Pleasing both authors and readers. A combination of short print articles and longer electronic ones may help us do this. *Br. Med. J.* **318**, 888–889
2. Markwell, J., and Brooks, D. W. (2003) "Link rot" limits the usefulness of web-based educational materials in biochemistry and molecular biology. *Biochem. Mol. Biol. Edu.* **31**, 69–72
3. Dellavalle, R. P., Hester, E. J., Heilig, L. F., Drake, A. L., Kuntzman, J. W., Graber, M., and Schilling, L. M. (2003) Going, Gone: Lost Internet References. *Science* **302**, 787–788
4. Koehler, W. (2002) Web page change and persistence—a four-year longitudinal study. *J. Am. Soc. Info. Sci. Tech.* **53**, 162–171
5. Koehler, W. (1999) An analysis of Web page and Web site constancy and permanence. *J. Am. Soc. Info. Sci. Tech.* **50**, 162–180
6. Institute for Scientific Information (ISI) Journal Citation Reports: JCR Science Edition (Institute for Scientific Information, Philadelphia 2003)
7. Ioannidis, J. P., Chew, P., and Lau, J. (2002) Standardized retrieval of side effects data for meta-analysis of safety outcomes. A feasibility study in acute sinusitis. *J. Clin. Epidemiol.* **55**, 619–626

Received for publication August 1, 2005.
Accepted for publication August 15, 2005.