

November 7, 2005

Link By Link

Beware Your Trail of Digital Fingerprints

By TOM ZELLER Jr.

IT hardly ranks in the annals of "gotcha!" but right-wing blogs were buzzing for at least a few days last week when an unsigned [Microsoft](#) Word document was circulated by the Democratic National Committee. The memo referred to the "anti-civil rights and anti-immigrant rulings" of [Samuel A. Alito Jr.](#), a federal appeals court judge who has been nominated to the Supreme Court by President Bush.

The stern criticisms of Judge Alito rubbed some commentators the wrong way (Chris Matthews of MSNBC called it "disgusting" last Monday). But whatever the memo's rhetorical pitch, right-leaning bloggers revealed that it contained a much more universal, if unintended, message: It pays to mind your metadata.

Technically, metadata is sort of the DNA of documents created with modern word-processing software. By default, it is automatically saved into the deep structure of a file, hidden from view, with information that can hint at authorship, times and dates of revisions (along with names of editors) and other tidbits that, while perhaps useful to those creating the document, might be better left unseen by the wider world.

(If you use Microsoft Word, open a document, go to the File menu and choose Properties. You should see some metadata. Third-party programs are available that will crack open even more.)

According to some technologists, including Dennis M. Kennedy, a lawyer and consultant based in St. Louis, ([denniskennedy.com](#)), metadata might include other bits of information like notes and questions rendered as "comments" within a document ("need to be more specific here," for example, or in the case of my editors, "eh??"), or the deletions and insertions logged by such features as "track changes" in Microsoft Word.

"If you take the time to educate yourself a little and know the issues," Mr. Kennedy said, "you can avoid problems pretty easily."

With the Alito memo - which was distributed on a not-for-attribution basis, with no authors named - the D.N.C. was a little sloppy.

Mike Krempasky, a conservative blogger at [RedState.org](#), mined the document's metadata and came up with juicy, code-cryptic tidbits like this (bold added for emphasis):

{lcub}o:Author>prendergastc{lcub}/o:Author{rcub}

Or this:

{lcub}o:Company>DNC{lcub}/o:Company{rcub}

"The technical wizards at the Democratic National Committee never got the 'don't forward Word documents' memo," Mr. Krempasky wrote, eventually identifying "prendergastc" as Chris Prendergast and "adlerd," which also showed up in the metadata, as Devorah Adler - both members of the D.N.C.

The metadata also coughed up a file creation date of July 7, 2005, which the detectives at RedState.org identified as being "just after O'Connor resigned."

None of these amounted to earth-shattering revelations, of course, but taken together they offered a level of detail into the Alito memo that the D.N.C. had not intended.

Josh Earnest, a spokesman for the Democratic committee, pointed out that the origins of the document were never really a secret, even if it was circulated as background material that was not intended to be sourced.

"Based on the fact that the D.N.C. was known to be circulating the document," Mr. Earnest said, "I'm not sure that RedState is breaking any news here."

Still, metadata and other document gaffes have tripped up other organizations, sometimes with more embarrassing results.

Just two weeks before the Alito memo, the United Nations issued a long-awaited report on Syria's suspected involvement in the assassination of Lebanon's former prime minister, Rafik Hariri. It was a damning report for Syria by any standard, but recipients of a version of the report that went out on Oct. 20 were able to track the editing changes, which included the deletion of names of officials allegedly involved in the plot, including the Syrian president's brother and brother-in-law.

A similar gaffe embarrassed the network software company SCO Group in 2004, when it filed suit against [DaimlerChrysler](#) for violations of their software agreement. A carelessly distributed Microsoft Word version of the suit revealed, among other things, that the company had spent a good deal of time aiming the suit at [Bank of America](#) instead. "It just sort of made it look like they were looking for the easiest target," Mr. Kennedy said.

At about the same time, California's attorney general, Bill Lockyer, floated a letter calling peer-to-peer file-sharing software - long the bane of the entertainment industry's interests - "a dangerous product." But a peek at the document's properties revealed that someone dubbed "stevensonv" had a hand in its creation.

Vans Stevenson, a senior vice president with the Motion Picture Association of America, said later that he had offered input on the document but had not written it.

"California AG Plays Sock Puppet to the MPAA," was one blogger's response.

The issue increasingly nags at the legal system, as lawyers become aware of the advantages of requesting discovery of the metadata buried in word-processed documents (or debate the ethics of scrubbing the metadata from a file before turning it over to the other side).

"If I get a piece of paper, all I see is a piece of paper," Mr. Kennedy said. "With an electronic document, there's potentially a lot more there." He noted that at a recent conference on electronic discovery, an Oregon lawyer complained that judges there tended to rebuff requests for the electronic versions of printed documents, saying the printed versions are enough.

But for most other instances - and certainly for cases like the Alito memo - the solutions are simple. Sort of.

Saving a copy of a document in "rich text format" (RTF), or as a simple text file first (options in the Save menu), and then converting it into the common "portable document format" (PDF) before circulating it is a good tack, Mr. Kennedy said. Still, some debate remains as to whether traces of metadata from word-processing programs like Microsoft Word are carried through to the PDF file.

For those who want to be extra safe, several third-party tools will scrub metadata and other information from documents, although with each new advance in software design, the number of potential pitfalls grows.

"It only gets more complicated," Mr. Kennedy said, making sure to point out that all kinds of documents - from spreadsheets to PowerPoint files - contain oodles of metadata. "It seems every time I turn around I run into something new."

Odds are that Derrick A. Max, the head of two business groups that favored President Bush's plans to privatize Social Security, wishes he could say the same.

On request last spring from the Senate Democratic Policy Committee, he e-mailed testimony on the topic. The unscrubbed Word document apparently included editing and advice from an associate commissioner of the Social Security Administration.

"The real scandal here," Mr. Max told The Los Angeles Times after Democrats expressed outrage over the White House's fingerprints on the testimony, "is that after 15 years of using Microsoft Word, I don't know how to turn off 'track changes.' "