

The End of History

Net Forecasts – Peter J. Sevcik

BCR Volume 30, Number 7

July 2000

On May 23, 2000, *The Boston Globe* published a critical review of George W. Bush's military record, which reported that during the early 1970s there were two years when he was in the Houston Air National Guard but never performed any of the duties associated with his commission. Obviously, this isn't the kind of news that makes a candidate's day.

I mention this story to illustrate a much larger issue, one that has nothing to do with how I feel about George Bush. Here's the real question: If all record-keeping moves onto the Web, will anyone be able to check facts? Won't it be easy to change old, embarrassing records?

This isn't as straightforward as you might think. On the one hand, it seems to be a good thing that history can be preserved, indexed and searched electronically, instead of having to rely on ink marks on paper which degrades with age, and plow through reams of often unorganized and unindexed paper records.

However, there's already been at least one instance where history deemed to be embarrassing was simply erased. In a letter to its readers, on June 1, 1998 the editors of *The New Republic* revealed that they had published 41 articles by an author named Stephen Glass, articles which they subsequently learned were based on fabrications.

The editors promptly dismissed Glass and issued an apology, but they took one additional step: They removed all of Glass' articles from the on-line version of the magazine. The article listings also disappeared from the archive Web pages entitled "Complete Table of Contents." If this magazine were only published on-line, the questionable articles would be completely gone, vanished, as they'd never been published.

It's important to note that not all publications react the same way. *The Boston Globe* endured a similar humiliation, involving articles by Patricia Smith. She also was fired, but her articles are still on *The Globe's* Web site. Interestingly, however, there is

no editorial footnote or other indication that the facts within the articles have come into question.

My informal poll of other journalists on how content errors are handled in on-line publications revealed a wide range of responses, ranging from "we have no policy" to "we would simply correct the error." Clearly this problem is too new for any consistent industry practice to have emerged.

There is No History

I first started thinking about this topic when I was trying to track the changes in the Internet and network vendors. The sites like Keynote Systems (www.keynote.com) that have measurements and/or indices of network performance only maintain a limited amount of data online for a very short period of time. When you ask to review – or even to buy – historical data, you quickly learn that it's not available. This makes the job of forecasting network traffic and technology a lot tougher. Just imagine trying to forecast the weather without historical meteorological data.

The networking industry is a leading indicator of how the rest of the economy will operate, and when companies promise to deliver something, a lot rides on their word – network planning and procurement cycles rely on information about future releases, as do stock prices. When major vendors (and wannabes) miss promised delivery dates, it usually sparks negative press, and credibility lost is hard to regain. That's why the good PR/marketing firms advise their clients to make sure that the story they give out publicly is the truth.

Today, however, Web-based PR seems to have forgotten that basic principle. Indeed, within just the past three months, I've seen two network start-ups totally change their respective stories. They've redone their Web sites and briefings with new content. Their original approach to the market has disappeared, as if it never existed.

History and Accountability Matter

As I was researching this column, I ran into a lot of skepticism about the need to be vigilant regarding the accuracy of information in the 'Net. Some

people believe that no one would dare change content on a Web site because it would be too embarrassing if they were caught.

Unfortunately, that's an increasingly naïve view. Indeed, while much has been written about the 'Net ushering in the end of privacy, it seems to also be ushering in the end of accountability.

I challenge any *BCR* reader to find the subtle changes to a product specification(s) on a Web site. How many people purchase on-line based on information only seen on-line? If Vendor A comes to realize that people are choosing Vendor B's product because of the technical data on B's Web site, how long do you think it will take for the content on Vendor A's Web site to change?

Remember, the game is changing. In the not too-distant future, when you pick up a prescription, instead of getting details about the medication on a tough to read, tightly-folded slip of paper, you'll be directed to the pharmaceutical company's Web site to learn about the drug's dosage, interactions and cautions for use. But will you be able to find out if the recommended dosage has changed over time – say, from 200 mg two years ago, to 100 mg last year and 50 mg today? Why is the dosage trending down? Good luck finding the answers.

Once the 'Net becomes the storage center for our important documents, imagine how easy it will be to fix a missed diagnosis on an old medical exam, to create a sudden change in zoning for the property next door, to revise taxes owed or product liability data and, yes, to add new hidden charges by a bank or phone company. (I realize the later already occurs, but at least you get a notice in the mail. In the future, you won't be able to decipher charges at all.)

Two Solutions to Two Problems

So, is anything to be done? Well, the outlines of solutions are beginning to be proposed. One approach is to create some kind of Web archive, and Brewster Kahle's Internet Archive Foundation (www.archive.org) is already working on this enormous task. As of mid June, the foundation had archived 14 terabytes of content, fast approaching the total textual information held by the Library of Congress. The size is impressive, but its use is

limited to historians who have access to special Unix-based search tools.

The archive is too big and complex to be mined by a consumer, but Alexa Internet (www.alexa.com) provides a free browser plug-in to help use the Internet Archive. The Alexa software is a collaborative tool by which the foundation gathers, manages and analyzes the multi-terabyte archive. When you click on a site, a special toolbar shows you whether an archived version of that site exists and, if it does, you can call up the archived version and make your own determination regarding any changes. Unfortunately, there is no tool that performs a "compare" function – to show if and where any changes have occurred.

It also includes *everything* on the Web page. For example, an old page that included a notation in the current weather will likely be different from today's weather. Alexa makes browsing more informative, but the average age of the content it is comparing today's site against is only about 2 months old. A key limitation is storage and so it's not clear whether, years from now, it will be able to serve a useful purpose to review historical – and huge -- monthly archives.

Another approach to the problem is to authenticate what's in the Web archives. A landmark bill is working its way through Congress that puts full legal backing behind digital signatures. That's good, but not enough; as more and more business moves exclusively to the Internet – i.e., no paper copies -- I hope that the security industry will look at how to "certify" Web sites. We will all benefit from a much larger market for digital certificate technology.

Imagine a service that creates and escrows a message digest or hash code of each Web page submitted by a client Web site. The Web site could post a "verify this page" button on its site, so any reader of the page could re-run the hash and check it against the escrowed code. The reply would state that from a date-certain the content of the page had not been modified. The service could charge to have the page placed in escrow and add another very small verification charge to the originating Web site per page that was checked.

Public standards for digital signature exist – e.g., RFC 1321 (a.k.a. MD5) – and are already widely used by the Unix and Linux communities; RSA or PGP can be used as well. The key is to get a neutral message-digest escrow company running, and then attract Web sites to join the verification service. The consumer would not need any special software or hardware; the site and its contents are being authenticated, not the user.

The escrow would only need to store the hash code and some clear-text source identification data. A hash can be stored in one-thousandth the space required to archive the total contents of the original Web page. This approach would verify sites that care about their content rather than the sites that are popular.

Such a service would not prevent removing content all together but it would make hiding changes much more difficult. In *The New Republic* example above, the articles might be removed, but the original Table of Contents for those issues would be viewable, and the hash of the contents would discover any missing text.

The 'Net is a massive change agent, but without historical records, the result is chaos rather than progress. It is time for the 'Net to grow up and become accountable.

Peter Sevcik is President of NetForecast in Waltham, MA, and is a leading authority on Internet traffic, performance and technology. He has contributed to the design of more than 100 networks, and led the project that divided the Arpanet into multiple networks in 1984, which was the beginning of today's Internet. He can be reached at peter@netforecast.com.