

Electronic Documents Are Different

William Badke

If a document is in electronic form, is it different from the same document in paper form?

At one level we'd have to say "no," because the content is the same regardless of its format. Format is thus irrelevant. Whether I read it in a book or off a screen, the message is the same.

But Marshall McLuhan, the Canadian techno-philosopher, argued that "The medium is the message," meaning that the format we use for communication carries as much of a message as the words themselves. This is certainly true for electronic documents. Electronic documents are not the same as print ones, even if the words are the same, for the following reasons, among others:

1. They can be distributed widely within a very short time period, thus making them far more contemporary than print.
2. They can be easily altered without detection.
3. Copies can look entirely different from one another in font, print size, spacing, color, etc., so electronic documents are more malleable than print.
4. If the document has hyperlinks to other document, the reader does not have to stay within the document itself. Do the hyperlinks thus become part of a larger "document" that includes the original text as well as the content of the hyperlinked pages?
5. If the document is hyperlinked within its own content (through bookmarks or a web of hyperlinked pages), the reader need not read it in order as is normally done with a print document. This disrupts the idea of a document as a linear progression of thought and may work against the interpretive guideline that every piece of data must have a context.

DEFINITION

Thus an electronic document disrupts the very meaning of the word “document.” Electronically, a “document” can be viewed from anywhere in the world at the same time via the Internet, can have its wording and its look changed at will without any sign left behind that there was an earlier version, and can encompass other documents as well as encourage reading out of order. This may be exciting (for example, we can hyperlink a document so that any possible problem or interest a reader may experience can be answered with the click of a mouse) but it carries dangers as well:

1. The fact that an electronic document can be created and flashed around the world in an instant may also mean that half-blown ideas can be shared as if they were the more well-formulated ideas of a print document.
2. The fact that a document can change both its wording and look without leaving a sign that there was an earlier version can lead to revisionism (rewriting history) of the worst sort. If words are vehicles to convey our identity, our best thoughts, our history, our essence, then the fact that those words can be changed, or the look of those words can be changed, can mean that evil people (or even misguided good ones) can revise our history or alter our identity. Memory being what it is, we will soon forget there ever was an earlier version of our understanding of ourselves. There are also legal ramifications in not being able to guarantee that a document has not had its wording altered.¹
3. The ability to hyperlink is also creating a loss of the sense of context. Data is becoming detached from the roots that give it meaning. Just as isolated words have no certain meaning unless put in a sentence, data in documents, if not read in context, loses its certainty, and true communication disappears. Hyperlinks provide us with sight-bites that lack clarity because we see each one in isolation. While the presence of footnotes and endnotes has always presented the problem of loss of linearity and detachment of data, hyperlinks have the power to draw the reader completely away from the original document.

¹ See Michael R. Overly, “The Admissibility of Electronic Documents,” Computer Forensics, Inc., 1997; reprinted from *The Los Angeles Lawyer Magazine*, June 1996. [Journal article]; available from <http://www.forensics.com/resources/admiss.htm>; accessed 18 January 2002. Ironically, the electronic version is no longer available on the Net. The above link is defunct.

All of this has led to renewed attempts both to preserve and define “documents” so that electronic data can be at least as stable as print while retaining its hyperlink capabilities.

COPYRIGHT

One approach to preservation of electronic data has been to rethink copyright for electronic documents. Copyright has traditionally controlled who can disseminate a document and in what form. It has also protected the text of the document by preventing it from being published in altered form as long as any of the content is the same as the original. A fascinating article by Charles C. Mann, “Who Will Own your Next Good Idea?” chronicles recent trends in the issue, from libertarians who have declared the whole notion of copyright dead, to legislators who want to put more teeth into copyright law.² Of particular interest is the concept of “licensing” documents, that is, delivering them in such a way that the user can have access only by agreeing to certain conditions imposed by the producer or vendor. If those conditions are violated, the user can be prosecuted as having breached a contract.

Copyright protection does have its place, particularly in trying to preserve the text and the look of a document, but there are powerful forces out there who are working very hard to end the “tyranny” of copyright (remember the first version of Napster)? In the end, copyright is only as useful as it can effectively prosecute infractions. Right now, prosecutions are not keeping up with breaches of copyright.

ARCHIVAL FORMATS

Other possible approaches to preservation may be more promising. One is “electronic watermarks,” a new and complex way of guaranteeing that the original text of a document, even its format, is not

² Charles C. Mann, “Who Will Own your Next Good Idea?” *Atlantic Monthly* 282, no. 3 (September 1998): 57-82.

altered. Any alteration would immediately be detected. This technology, while still in development, looks promising.³

An older but possibly sturdier solution may be to produce documents as static images rather than text, using the Adobe PDF format or Xerox's DigiPaper. A very promising development in this regard is PDF-Archive (PDF/A), a rigorous new PDF standard for archival documents, due in 2005.⁴ Such documents could not be altered in themselves but would have to be copied to be altered, so that originals would remain untouched.

But the electronic medium has created a new problem - documents (particularly social scientific and scientific documents) are now being produced with links to videos, tables, notes, and so on that are integral to these documents but are located at a different URL. Page image solutions would have to accommodate linking, which again raises the question of whether the document itself or the document plus that to which it links is the actual document. What exactly do we need to preserve? If we save the original document but its links go out of date, have we lost something that was crucial to preserve? Should we "cache" everything to which a document links and preserve it along with the document?

DURABILITY

Another crucial problem is the durability of electronic documents. Web pages come and go. Sometimes you can retrieve a lost web page by pulling up the cached version in Google, but eventually it will go forever. Even documents on hard drives or CD-ROMS, DVDs etc. are at risk because drives crash and formats change. While paper has its own risks (fire, flood, deterioration), it tends to be more durable

³ Purdue University. "Purdue Team Develops Watermark to Protect Electronic Documents." 27 April 2001. [News release]; available from <http://www.sciencedaily.com/releases/2001/04/010427071702.htm>, Accessed 16 November 2004.

⁴ For ongoing news on PDF/A, monitor http://www.aiim.org/pdf_a/ (accessed 16 November 2004). See also Dan Huttenlocher, "On DigiPaper and the Dissemination of Electronic Documents," *D-Lib Magazine* 6 (January 2000). [Journal article]; available from <http://www.dlib.org/dlib/january00/moll/01moll.html>, accessed 16 November 2004; Nikki Swartz, "Will PDF Prevent a Digital Dark Age?" *Information Management Journal*, 37, no. 3 (May/June 2003): 13.

than an electronic document. There seems to be little effort out there to preserve electronic documents in a systematic way. A possible proposed solution to disappearing web pages is the URN (Universal Resource Name) as a tag that identifies a document regardless of its location. By searching on the URN, a document in a new location can easily be found.⁵ Recently the U.S. government has considered utilizing URNs for all of its web pages so that changing departments and information will not be lost to the searchers who need them.

REVISION

Finally, those who have access to the production and copyright of a document can themselves alter it in significant ways. Every time you see a note at the bottom of a web page indicating, “Last revised June 24, 2001,” you know that the document you’re looking at is not the same as the document that existed June 23, 2001. [In fact, the document you are reading has been revised]. In the book world, revisions are marked by new editions, and we can get access to the old edition to see what has changed (though it was Britannica that years ago initiated a program of continuous revision, that would bring changes, often unidentified, every time a new printing was done.) Now, with electronic documents, revision can happen at will. While this keeps these documents relevant, it can also change, without our knowledge, the information we rely on. Thus power is put into the hands of the writer who can reshape knowledge simply by revising his/her documents.

One fascinating ramification of the capabilities of electronic revision is the development of collaborative document production. Collaborative research documents can be created by a process of several scholars, perhaps separated geographically, contributing to and revising a communal work. Murphy et al. describe the phenomenon this way:

Since the advent of e-mail and fax, researchers across distances have used telecommunications to write documents together – getting feedback from co-researchers and incorporating those ideas into the

⁵ See “Digital Object Identifiers and Metadata for Electronic Documents.” National Computing Center. Originally at: <http://www.ncc.co.uk/standards/projects/doim.html> (18 January 2002). By a strange twist of irony, this document was removed from this site as of mid-March 2002 and reposted to <http://www.ncc.co.uk/research/projects/doim.cfm>. Only by careful sleuthing was this author able to find it again. Thanks to Lech Lee from Hong Kong for alerting me to the loss of this document.

manuscript. Eventually, through multiple iterations, the document becomes a final product.⁶

Another growing trend is the Wiki, a web environment in which users can alter or add to web information at will. Wikipedia is the most prominent Wiki example – an online reference tool created by users without the services of an editor. Anyone can write an article or revise one at will, thus providing the ultimate in democratization of information dissemination.⁷ All the challenges, of course, of quality and bias are enhanced in such an environment (resulting, just before the 2004 election, in the article on George Bush having to be pulled from Wikipedia for a short time due to the number of dramatic revisions being done).

CONCLUSION

Bottom line: Electronic documents are different. On the one hand, they can open new worlds and new ways of thinking, providing opportunities for currency and the democratization of knowledge such as we have never before experienced. The popularity of the Net and of full text academic databases makes it clear that electronic information will grow dramatically in size and use over the next decades.

But electronic documents, as delightfully promising as they are, also present serious problems to the stability of information. Coming years must bring solutions to the abuse of such documents so that they may have the same, or nearly the same, stability as print documents.

Revised version: November 16, 2004

⁶ Karen L. Murphy et al. "Online Collaborative Documents for Research and Coursework," *Tech Trends* 48, no. 3 (2004): 40.

⁷ *Wikipedia: The Free Encyclopedia*. http://en.wikipedia.org/wiki/Main_Page. Accessed November 16, 2004.