Slaying the Paper Dragon

THIS MONTH MARKS THE START OF MY NEW COL-umn about digital life. Most of *Technology Review* focuses on the companies and people driving innovation, and how those innovations will affect the world. But here, I'll explore how technology can be incorporated into daily lives. This column will be part product review, part how-to guide, part reflection on the impact of technology, and hopefully a lot of fun.

(Readers can find my "Net Effect" column every month on technologyreview.com.)

One of my big projects this past summer was cleaning out the basement. Like many people's, I suspect, my basement was filled with file boxes containing important documents and papers. Or at least, documents and papers that I once thought were important. There were old bank statements and telephone bills, letters my mother sent me at summer camp, my daughter's kindergarten art projects. As a writer, I had also saved research materials collected over a decade's work on books and magazine articles.

It's an old observation that the digital revolution has paradoxically flooded us with paper. But I finally decided that I had had enough and set about to liberate the data from all those dead tree shavings.

So I'm scanning those papers and putting many of them on the Internet. Two things pushed me into action. One is guilt: I have accumulated many paper files on the theory that they might be useful to somebody, someday. But as long as these documents are trapped in my basement, nobody knows they exist. The other is that the technology for turning printed matter into digital has become powerful enough and easy enough to use that I felt I had no excuse not to give it a shot.

The first part of my task—the scanning—hasn't been that hard. My Hewlett-Packard printer/scanner/fax/copier takes a stack of paper and automatically scans its content into Adobe Acrobat files. (To

> I decided "enough," and **set out to liberate paper data** from my basement boxes: I'm scanning my files and putting them online.

be honest, I only scanned the first hundred pages; then I hired a high-school student to do the rest.) It was then simple to put the files online using standard Web publishing tools.

But the problem with these scans is they're just pictures of the original documents. They look great, but Google—which searches the Internet for words and phrases—will never index them. Although I could use optical character recognition software to convert the images into searchable text, that would take a lot of time and introduce errors. Instead, I have written a few words to describe each document—something like "Social Security Report, *Privacy Journal*, 2000"—and then put those words and a link to the scanned document on my Web site. Only about a thousand people have downloaded that scan so far. Still, that's a thousand people who probably wouldn't have gotten that report at all otherwise.

Other stuff is not for public consumption, so I'm storing those files securely on my server. Two years ago I bought a digital camera and a copy stand with floodlights on the side and a camera bracket. Since then, I've been photographing my daughter's creations rather than archiving them downstairs. Three drawers of paper artwork have now been captured in 100 digital photographs.

I'm not alone in creating a vast personal digital archive. A friend in Colorado is digitizing his 35-millimeter slides with a slide scanner he bought on eBay; I have dibs on the scanner when he finishes.

What's making these archives possible is the huge capacity of today's disk drives: the scans of the reference materials from my book *Database Nation* take up nearly 300 megabytes. That was a lot of space when I wrote it back in 1999, but my digital camera today has more storage in its flash memory card.

But if you start creating your own digital archive, you'll discover that digitizing information and entering keywords for Internet search engines is only half the task. You also have to organize digital files so that you can find what you've archived years from now. This complicated job requires, in addition to making backups, a taxonomy that allows you to enlarge and extend your database over decades. Yet another problem, for those making information public, is securing permission from copyright holders to put the data online. You can buy specialty software that fulfills many of these tasks. Unfortunately, these programs store data in proprietary formats. Since I hope to keep my data for 40 or 50 years, that constraint is bound to create hassles down the road: who knows what formats will be supported by systems then in use?

Clearly, this paper escape is still too cumbersome for most people. But if you have both a lot of knowledge about computers and a willingness to devote a good chunk of time to solving problems, you will find the challenge worth tackling.

Now, if I could just digitize those boxes of clothes in my basement. Ⅲ

Simson Garfinkel is an incurable gadgeteer, an entrepreneur, and the author of 12 books on information technology and its impact.