

[Click here to Ask Jeeves.](#)



**HOTWIRED**

[Wired News](#)

[Webmonkey](#)

[RGB Gallery](#)

[Animation](#)

[Archives](#)

[Wired Mag](#)

[Suck.com](#)

Search:  for

# Let My Data Go!

## Freeing file formats would free us all from software monopolies

by [Simson Garfinkel](#)

6 January 1998

An important issue is being overlooked in the fight between Microsoft and the rest of the software industry. Today, the industry and the US Justice Department seem focused on preserving competition in the world of Internet applications, like Netscape Navigator and Sun's Java language. I'm more interested in seeing competition return to the world of desktop productivity applications. I want to see real competition (and, as a result, real improvement) for programs like Microsoft Word, Microsoft Excel, and Intuit's Quicken.

I propose an improvement that focuses on an often-invisible aspect of the software market - file formats.

Some people use Microsoft Word because they think it's the best word-processing program around. But others use Word because their office, company, or university employs the software to the exclusion of all other word-processing software. Says one disgruntled Word user: "I need to be able to send files to people on floppies or attached to email, and I cannot expect clients to translate files. Everyone in my company uses MS Word, so all templates and letterheads are set up for Word."

Other word-processing programs have limited support for reading and writing files created by Microsoft Word; these programs can't do a perfect job. WordPerfect and ClarisWorks lose important formatting information when they import Word files. The reason is that the way in which Word stores information on files on a computer's hard disk - the program's file format - is a corporate secret.

Programmers have figured out how to "crack" some of the Microsoft Word file format, but not all of it. And Word is not the only program with a secret file format. The formats used by nearly every mainstream computer program are undocumented. Whether you are using Microsoft Excel, Intuit's Quicken, or an Oracle relational database, you have no way of knowing how your program is storing data on your computer.

There are many good reasons for forcing companies to reveal these details and document how

their programs store information. Companies argue for file-format secrecy in order to protect their intellectual property. I think this argument is as hollow as a tobacco company saying it needs to keep its cigarette additives secret for competitive reasons. The company's right to propriety should be trumped by the public's right to know.

Forcing companies to document their file formats would remove the single largest barrier to entry that start-ups face when trying to enter an established market: compatibility with the installed base. There is little hope that some new company would try to challenge Microsoft Word with new desktop word-processing software, because that company's program almost certainly couldn't read all of the subtle formatting commands that Microsoft Word files contain. So it is unlikely that, no matter how revolutionary its features, it could capture a significant market share.

Beyond creating new opportunities for competition, forcing companies to disclose their file formats would improve the quality of software.

Most file formats in use today are not things of beauty. They are often ugly hacks created by programmers who are hard pressed for time. Some companies do not even have a formal review process to decide if their file formats are up to the task at hand. Some companies don't even have file formats properly documented.

The problem is that a bad file format can have deleterious effects for the user many months or many years in the future. For example, an accounting program may store its dates with 2-digit years, making it susceptible to the year-2000 glitch. Or there may be a conceptual error in the program's design, in which two pieces of critical information are stored in the very same location. These sorts of problems can be very difficult or even impossible to detect by testing, but they are often apparent upon inspection of a file-format description.

If companies were required to disclose their file formats, there is a good chance that the mere act of creating this documentation would force the companies to create a better product. This is analogous to the Food and Drug Administration forcing companies to print food labels with a list of ingredients. Improve the disclosure of information, and the market will work to improve the products.

Documented file formats would also help launch a new generation of after-market programs - programs that could read the data files created by other programs. A few years ago, I was working at a magazine that wanted to take articles from a few dozen back issues and put them on the Web. But we couldn't, because all of the articles had been formatted using QuarkXPress, and we didn't know how to decode the Quark file format. Our data was being held hostage, and there was nothing we could do.

This data-hostage dilemma is a life-and-death issue for smaller companies and their customers. Back in the 1980s, I used a personal finance program called Dollars and Sense. When the company went out of business, my data was trapped. There was nothing I could do to get it out. As a result, I've been wary ever since of taking a chance on a software start-up for fear that more information could be lost to the abyss. I simply can't afford to lose access to my own information.

How would a law or regulation requiring file-format disclosure work? My favored approach would be to require companies to distribute the information with their programs. For example, the details

could be put into a "help" file or included on a CD-ROM. The expense would be minimal. Although some companies might elect to create detailed written specifications, others could satisfy the requirement by simply distributing the file-reading source code from their own programs. Such source code typically represents less than 1 percent of the code in a major application.

Companies could face heavy penalties for refusing to comply. A regulatory agency could require that the products be recalled, the way unsafe children's toys are pulled from the market. But I think that a better tactic would be to simply deny copyright protection to programs that do not reveal their file formats.

Ultimately, competition in the world of computer software depends on equal access to information - for consumers and competitors alike. Today too much of our nation's data is being held hostage in proprietary file formats. Congress should take notice and craft new legislation designed to let our data go.

. . . .

Would freeing file formats really free up the industry? Talk about it, in [Threads](#).

More from Simson Garfinkel:

[Adoption](#) policy and the Net

50 ways to [crash](#) the Net

Garfinkel on [info-espionage](#) online

Electronic [border-control](#)

Debating the future of [privacy](#)

The proposed [SET](#) protocol is totally unnecessary

Browse the [Garfinkel](#) archive

Email [Synapse](#)

[synapse](#)

[Wired News](#) | [Wired Magazine](#) | [HotWired](#) | [Webmonkey](#)  
[RGB Gallery](#) | [Animation Express](#) | [Web 101](#) | [Suck.com](#)

[Work at Wired Digital](#) | [Advertise with us](#) | [About Wired Digital](#) | [Our Privacy Policy](#)

[Copyright](#) © 1994-98 Wired Digital Inc. All rights reserved.