

Why data should be in open formats

Bas Wijnen

14th May 2002

Abstract

This document describes the problems of using non-open data formats that do not arise when using open formats instead. It hopes to convince the reader to use open formats and tell others to do likewise.

1 the writer

The writer wants a nice user interface while writing the document. This does not usually make much difference, since good user interfaces exist for both open and non-open formats.

Furthermore, the writer wants that the receiver can read the data. If in any way possible, the reader should also have the data presented as it was written. Greek letters should still be there, formulae should look as intended by the author, etc. While the most known non-open format (Microsoft Word) does not comply with this demand, it is not a reason to discard non-open formats, since they can very well comply (Word Perfect files comply, for example).

The writer want to give his/her license to his data. It is very well thinkable that in the future, non-open formats come with restrictions on the license you may place over you own data. While this is very disturbing, it is also not a reason to discard non-open formats completely, only to be cautious with the composing software to use.

2 the receiver

Just as the writer does, the receiver will first of all receive the data in a form which is readable and which presents the data in the form it was written. The receiver wants this without any restrictions. To read Microsoft Word documents, for example, a word viewer is needed. Microsoft does provide it, but only for Microsoft Windows users. This means that in order to read Microsoft Word as it was intended, the receiver must accept the license of Microsoft Windows (and pay money to be allowed to use it.) This is a valid reason for not accepting Microsoft Word documents. But it still does not give a reason for discarding all non-open formats.

In many cases, the receiver will want to modify the data, to send the modified version back, or to other people. Sometimes this is the very reason that the data was sent. In this case, a non-open format (which generally can only be edited with the owner's program, which imposes a license on its users) implies that the receiver must sign a license with a company that he/she probably doesn't want anything to do with. In the cases where this happens, however, the receiver usually has signed this license already, so this is not a problem. It does rise problems for society, however, which I shall treat in the next section.

3 society

What does society have to do with me sending a file of data to a friend? Society doesn't even know. However, if you use a certain data format to send something to a friend and you didn't ask that friend before if he/she is capable of reading the format, then this friend will conclude that it

is normal to assume that everyone can read the format you sent. Therefore, the friend may start sending files in the same format to other people. If this is an open format, there would be no problem. Since the format is open, people can indeed read it without any limitation. But with a non-open format, things are different, as described above.

Moreover, if people keep sending each other non-open formats, they will start considering it normal to assume that everyone can not only read, but also write data in that format. The result is that it may happen that people can only apply for a job if they send their application letter in a non-open format (which can only be done properly if the applicant uses the software from the format owner and thus has signed the license.) This is very bad for society.

Furthermore, if the format owner decides they want more money, they can “upgrade” their program and people with the old version will not be able to read the new data. Especially when Microsoft is concerned, this is a problem, since they will not even produce a new reader for the old platform (they usually “upgrade” the operating system at the same time and everybody thus needs the new version to be able to read data they receive from friends.) This practise of abusing their monopoly is not at all unknown in the history of Microsoft and it is likely to happen again, and again, and again. Until people stop using their products. Or at least stop using their data formats.

4 other people involved

There is only one other party involved and that is the owner of the non-open format. As described in the previous section, there are economical reasons for the owner to keep the format closed. Also, it is financially attractive for them to harm the public by demanding all kind of things in their license. While any non-open format introduces such dangers, Microsoft is a special case. Because of their monopoly in the operating system market, they can demand anything at all. They have demanded quite a lot already and got it all (which is why WordPerfect went out of business). While the owner wants everyone to use their data format, it would be much better for the public to use an open format.

5 conclusion

For the private citizen, it is best to not only use open formats, but also to advise others to do so. While owners of formats, most notably Microsoft, will tell many things in commercials which may seem nice, the fact that the format is non-open harms society too much to even consider using it.