

Why not word ?

RFC1149

(<http://www.rfc1149.net/documents/whynotword.html.en>)

French version available at <http://www.rfc1149.net/documents/whynotword.html.fr>

Summary

This short document explains why you should not use proprietary formats such as Microsoft Word or Microsoft Excel to distribute information.

Work in progress

When you need to work with co-workers on one document, you have first to agree on an exchange format which allows anyone in the team to make the necessary modifications. It means that all the co-workers must have one software which can understand and write the chosen format.

For example, if all the co-workers do buy a copy of Microsoft Windows or Mac OS and do buy a copy of Microsoft Office, it is possible for them to chose the Microsoft Word or Microsoft Excel proprietary formats. As long as all the team members agree, there is no problem with that (except a reliability one, which won't be handled in this document).

Distribution

Distribution is another story. When you want to distribute a document to a group of people, you cannot assume that those people do own the same software as you, and even if they do, you cannot assume that they do have the same version. Sending them a document in a proprietary format actually means that you want them to buy the same software as you did to be able to read your document.

Fortunately, there are other means to distribute a document: you can use standard formats. For example, the HTML format has been standardized by the W3 consortium. It is possible to write a software that understands this format on any platform.

Another simple format is the plain text format: when a document contains text only, bells and whistles are not necessary and do not make the content any clearer. Moreover, you can be sure that your readers will actually be able to read this format, as long as you use a common character set.

Confidentiality

Work in progress documents contain a lot of useful information. For example, most word processors allow you to retrieve previous versions of the document. In a spreadsheet program, you can see and change formulas, and hidden columns may contain other data such as the discount rate or your gross margin.

When you distribute a commercial proposal or a text that you have reworked, you probably don't want your receiver to access all this information. Your client would know that you hesitated about the discount rate to apply, and that you didn't offer him the greatest one.

When you distribute your documents in an adequate format such as HTML, PDF or PostScript, you control exactly what you send to your readers. Hidden data are simply not included in the final document, and formulas are not visible either.

Documents archiving

Documents archiving is yet another problem. Let's assume that you store your documents on a CDROM using Microsoft Word proprietary format; five years later, you want to reuse them.

Several solutions exist:

1. You still use Microsoft Word as your favourite text processor (of course, you still pay the license fees). In this case, you have to hope that the current version will be able to reread your saved files (only Microsoft can choose whether this will be possible or not).

2. You still use Microsoft Word as your favourite text processor but the current version is no longer able to read your saved copy. You have to reinstall an older copy of Microsoft Word, if you still own it and if the current Microsoft Windows version supports it. Or you can find an older machine and reinstall the older Microsoft Windows and the older Microsoft Word on it. However, if you have used an update path for Microsoft Windows and Microsoft Word, then you lost the rights to use your older versions.

If you had chosen to save your documents in an open and structured format (such as XML), you can be certain that you will be able to recover them easily, even in case of a major technology change. If you cannot do it yourself, you can hire a programmer who will be able to write such a filter without violating any license.

What about security?

Using proprietary software such as Microsoft Word creates security issues:

* Microsoft Word or Microsoft Excel documents may contain macro-commands, which will be executed when the document is open. A user can forbid the use of macro-commands, but in this case the document may not be seen in its entirety, as macro-commands may have been used unmaliciously. For this reason, more and more people refuse to open electronic mails if they contain a Microsoft Word attached file.

* Some releases of Microsoft Word use a fast save that dumps the program memory onto the disk. By exploring a Microsoft Word document, one can sometimes find sentences that have been since deleted or coming from a document that had been previously opened by the sender.

The humanity and history point of view

Another important point deserves our attention: free access to the information for everyone. As we are trying to give net access to the poorest people (be they poor people in Northern rich countries or poor people in Southern poor countries), it seems strange to force them to accept constraints they cannot satisfy:

* Using proprietary formats forces them to acquire the original software or a clone of it.

* Proprietary software may require hardware (e.g., Pentium III) out of sight of the poorest people.

* Proprietary software may require proprietary operating systems; for example, you have to buy Microsoft Windows or Mac OS to run Microsoft Word.

Moreover, who gives you any guarantee that in a few years those companies whose software use proprietary formats to save your documents will still exist ? In this case, who gives you any guarantee that those documents can still be read and worked on ?

To summarize

The conclusion is simple: try, as much as possible, to send your document in text, HTML, PDF or PostScript format.

Samuel Tardieu (22 October 2001, last revised on 23 October 2002)

Thanks to Nicolas Berdugo, Sébastien Blondeel, Brent Frère and Raphaël Rousseau for their suggestions and remarks.