

# MS-Word is *Not* a document exchange format\*

Jeff Goldberg

June 15, 2005  
version 1.35

Typically you are getting this because you sent someone an email message using MS-Word or some other operating system or text-processing specific attachment. Alternatively, you may have placed MS-Word files on the web as the only means for getting at the document content.

## Contents

<b>1</b>	<b>What's wrong with sending MS-Word files?</b>	<b>2</b>
1.1	Requires proprietary software . . . . .	2
1.2	Version problems . . . . .	2
1.3	Proprietary data format . . . . .	2
1.4	Viruses and security . . . . .	3
1.5	Size . . . . .	3
1.6	Prior version info . . . . .	3
1.7	Typically attached "wrong" to email . . . . .	5
1.8	Word is not device independent . . . . .	5
1.9	Word isn't even good at what it is designed for . . . . .	5
<b>2</b>	<b>Alternatives</b>	<b>5</b>
<b>3</b>	<b>Where MS-Word is appropriate</b>	<b>6</b>
<b>4</b>	<b>Response to the "it's the emergent standard" refrain</b>	<b>6</b>

---

\*See section 5.5 for copyright notice.

<b>5</b>	<b>History and related documents</b>	<b>8</b>
5.1	Similar documents . . . . .	8
5.2	Rants about MS-Word . . . . .	10
5.3	Reaction so far . . . . .	10
5.4	How you can help . . . . .	11
5.5	About this document and copyright notice . . . . .	11
5.6	Shameless plug . . . . .	12
5.7	Acknowledgement . . . . .	12

## 1 What’s wrong with sending MS-Word files?

### 1.1 Requires proprietary software

You are basically assuming that everyone has on their desktop the same software that you have. That often goes against the spirit of the Internet which is supposed to be about inter-operability of heterogeneous systems. That fact that one “persistently predatory monopoly”<sup>1</sup> attempts to subvert that goal, doesn’t mean that you should go along with it.

Someone who sends me such mail is perfectly welcome to purchase for me a machine and software specifically so that I can read mail in that proprietary system. But I will still have the inconvenience of having to forward the file to a system I wouldn’t normally use.

### 1.2 Version problems

Even for those who chose to use MS-Word, there are compatibility problems between various versions. Foreshadowing the next topic, it appears that Microsoft is unwilling to provide fixes for very substantial security problems in older versions. An article on CNN’s website (September 13, 2002) reports such an instance.

### 1.3 Proprietary data format

The above two problems are closely tied to the question of proprietary data formats. When you store your work in MS-Word format, you are betting that you will always have access to some licensed software that will be able

---

<sup>1</sup>In the words of a U.S. federal judge.

to read that format. The Open Data Format Initiative has more information on what is wrong with closed formats.

## **1.4 Viruses and security**

MS-Word allows full macro-scripting. It is now the most common carrier for viruses. What this means is that embedded within a Word file can be a program which runs silently (or otherwise) on the recipient's computer whenever they view the file. Are you happy with letting other people run programs on your machine?

In one instance that I know of, a substantial portion of an MBA graduating class sent out résumés with a Word macro virus. I don't think that this helped their job prospects. But the particular business school had an official MS-Word policy.

## **1.5 Size**

Often what would be just a few kilobytes of plain text is hundreds of kilobytes as a Word file. I find it interesting that MS-file browsers and emailers don't make it obvious to the sender how large particular files are.

## **1.6 Prior version info**

Because of Word's system of doing version control, it is possible that recipients may see prior drafts of your document (which may contain confidential information).

I've heard a number of "friend of a friend" stories about this sort of thing. In one case, a potential customer was given a quote for some product, and the quote was sent in an MS-Word file. When the customer viewed the version history, they found that a previous version of the document had been used for a quote to other customers, with much lower numbers. But since initially writing this, I have heard a number of first hand accounts. Some of which are below. Since I almost never read MS-Word documents sent to me, I will have to rely on the accounts of others.

Probably one of the most spectacular instances of information inadvertently leaked because someone (the British Prime Minister's office) used MS-Word for document exchange is described in an article by Richard M. Smith,

*Microsoft Word bytes Tony Blair in the butt.* The edit history of the “February dossier” has become a matter of contention to say the least. Smith’s article provides links and details.

Other, more mundane, accounts of meta-data leaking from MS-Word documents follow.

In a Usenet news article, Alan Frame describes some of his experiences with this

In the past, I’ve received MS Word documents from an agency, describing a job vacancy where they’ve refused to name the client – lo and behold, the document properties reveals all.

And also

Indeed, I’ve also seen an internal business proposal which appears to have originated at the supplier that the proponent was err, proposing.

I have also received word from others saying,

This regularly happens to me because I deal with public relations companies who always use the very latest spiffy version of Word and Powerpoint and seem to be totally unaware that not everyone does the same.

Normally I junk these docs, but if I need them I view them . . . and often see where corrections have been made. . .

I have never seen anything really sensitive as a result of this, probably because most press releases aren’t on very sensitive subjects. Usually I see comments like “CLAIRE: should we describe what the possible treatment options might be?”, plus minor word-changes. But I live in hope.

Charles Wankel posted a message concerning this to the E-Media list of the Academy of Management saying,

I received a paper for an effort that I was an editor for from someone who had used a ghostwriter. The ghostwriter had had embedded her name in such a way that when I looked at the document in a source view I could see it with the dates that wrote, edited, and re-edited drafts of the document.

## 1.7 Typically attached “wrong” to email

While this is not strictly speaking a problem with MS-Word files, it is a related problem. People and systems that think that it is right to just send such things, seem to think that it is OK to send everything with the MIME Content-type of `application/octet-stream` and let the recipient work things out from the filename info that is also sent. That is a violation of the intent of the MIME standards, and indicates broken design for exchange of information.

## 1.8 Word is not device independent

I have been told that MS-Word documents will format differently depending on the specifics of the printer. This is not merely issues of printer resolution or color depth, but the actual formatting of the document will differ. I was surprised to learn this. I had assumed that Word was “What You See Is What You Get”, but it appears that I was mistaken about that. So it won’t even achieve the goal of ensuring that your recipient sees things with all the formatting you see things with even if the recipient also uses MS-Word.

## 1.9 Word isn’t even good at what it is designed for

As an aside, I feel that MS-Word produces probably the worst output and is the slowest and most tedious to work in of any document preparation system in serious use I’ve seen in the past 15 years. I find it remarkable that when people are presented a choice between a structural mark-up system (what you mean is what get) versus a visual mark-up system (what you see is all you get) people opt for the latter. For more on this point see section 5.2. Note that the argument that MS-Word is an inappropriate exchange format is independent of this point about its quality as a document preparation system.

## 2 Alternatives

When talking about things sent by email it is important to distinguish between document exchange and message exchange. Message exchange is typically what one does by email. Making announcements or participating in a discussion, and many of the other things we typically do with email. For

these plain text is the only reasonable thing. It is the safest, most portable and by far the most compact. It allows responses quoting portions, and has none of the dangers mentioned above. The small added value of the formatting information isn't worth all of the problems.

If you absolutely *need* to present the formatting information for document exchange, then use a page description language like PDF.

Also consider using (standards compliant) HTML. Please note that I am not in any way advocating the use of HTML in ordinary email. It is grossly inappropriate for that for reasons that are beyond the scope of this document.

In earlier versions of this document, I listed RTF (Rich Text Format) as a more standards based way of exchanging word-processor documents. I have been corrected on that point innumerable times. RTF is little better than MS-Word format itself. It is a *little* better, but it shares all of the problems as MS-Word. Although RTF was advertised as a document exchange format, it never lived up to that. It appears to have varying features, and the various version of RTF that Microsoft products create have elements which only Microsoft Products can read. Note that this is not because MS-Word is a better product, but because Microsoft keeps elements of what it considers to be RTF secret.

### **3 Where MS-Word is appropriate**

MS-Word is appropriate for document exchange among co-authors of a document who are all developing it and have agreed before hand to use MS-Word. If you have been referred to the document you are now reading, then the person who referred you to it probably doesn't consider themselves party to such an agreement, and having sent them an MS-Word document is inappropriate.

### **4 Response to the “it’s the emergent standard” refrain**

Several people have responded with sophisticated “network analysis” essays about MS-Word being a *de facto* standard, and pointing out that even if the standard isn't the optimal one, it is better to go along with the standard anyway. My counter argument is two-fold:

1. Whether or not the argument about emergent standard holds for authorship (eg, “I use Word because it is what my potential co-authors use”) has little bearing on what you use for document exchange. I use  $\text{\LaTeX}$  for document preparation, but I distribute them as PDF.<sup>2</sup> So there *may* be an argument for using MS-Word even though it is inferior to other options, but that in no way suggests that MS-Word should be used for document exchange.
2. The second argument is an ethical one, and I start with an analogy.

Over the past few years it has become fashionable in the US to drive some form of truck as a primary commuting/errands vehicle. There are many issues regarding that fashion, but for this analogy I would like to focus on two of them. When two vehicles collide the occupants of the lighter one are far more likely to suffer injury than they would if they had collided with an equally light vehicle. So when someone drives a truck, they are putting those in normal sized vehicles at an extra risk. The second property is similar. The headlights of the trucks are much higher off the ground than those of cars. Driving a car at night with one of these trucks close behind you is extremely annoying and possibly dangerous. In both of these cases, the drivers of the trucks don't experience the disadvantage of others driving trucks. In the first case, they too are in heavy vehicles, and in the second the driver is high enough off the ground to not be impaired by the headlights of other trucks.

By the logic of the “emergent standard” advocates, the only way to deal with the truck problems I've described is to switch to driving a truck oneself. The emergent standard argument might have some validity if the standards were arbitrary, but if some are particularly destructive to community as a whole, they should be opposed. Use of MS-Word for document exchange is simply bad network citizenship. Paraphrasing Juhapekka Tolvanen: using MS-Word is like smoking; using it for document exchange is like blowing your smoke in everyone else's face.

---

<sup>2</sup>Using  $\text{\LaTeX}$  does have exactly the cost described by those who raise the “*de facto* standard” argument: I find myself limited in co-authors to a subset of clueful, intelligent and network cooperative individuals.

3. There is a third argument, closely related to the second: Do you want to be part of Microsoft's marketing effort?

## 5 History and related documents

### 5.1 Similar documents

When I first wrote the first version of this document in March, 2001, it was because I not only was fed up with people sending me unwanted MS-Word documents, but because I was tired of explaining repeatedly why I objected to them. I wrote this to be part of a canned response.

Being remarkably lazy, I didn't want to investigate and write this up if someone else had already written something. So I did a little bit of searching for documents like this. I knew from personal communication that while I am in a minority there is a substantial minority which feels exactly the same way. I expected that someone would have already written something like this document.

I didn't find any when I looked, but clearly I didn't look carefully enough. I have since been informed of others that I've missed. I list them here, along with some which were written after my document.

**plaintext: In praise of practical e-mail hygiene** This is Martin Vermeer's essay. It covers the same points as mine but goes deeper into trying to persuade people to be better network citizens.

<http://www.netby.dk/0est/Europa-Alle/vermeer/plain.html>

**We can put an end to Word attachments** This is an article by Richard M. Stallman advocating efforts like mine to discourage people from sending MS-Word documents. The article itself is aimed at those who already know that Word attachments are wrong.

<http://www.gnu.org/philosophy/no-word-attachments.html>

**Sincere Choice** This is the home page of the Sincere Choice platform who say "We believe that there should be a fair, competitive market for computer software, both proprietary and Open Source."

<http://sincerechoice.com/>



The Sincere Choice principles of open standards and interoperability underly much of what has been stated here.

[http://sincerechoice.com/Principles/Open\\_Standards.html](http://sincerechoice.com/Principles/Open_Standards.html)

[http://sincerechoice.com/Principles/Choice\\_Through\\_Interoperability.html](http://sincerechoice.com/Principles/Choice_Through_Interoperability.html)

**Open Data Format Initiative** This is an attempt to encourage software companies to fully document the formats of their data files. To paraphrase earlier words of the founder of this initiative, if you own the data in the PowerPoint presentation you created, why should you need a license from Microsoft to get at your presentation?

<http://odfi.org/>

**Miksi on typerää postittaa sähköpostin...** As you can see, this detailed essay and analysis by Juhapekka Tolvanen is in Finnish. I don't read that language, but there are some useful links from that. He comes up with a very useful analogy, which I will rephrase more harshly: Using MS-Word is like smoking; emailing those files is like blowing smoke into other people's faces.

<http://www.cc.jyu.fi/~juhtolv/mswordmail.html>

**MS-Word? nom obrigado** A similar document to mine, available in Portuguese and Galician, by Ramón Flores d'as Seixas. While this document is based on the others listed here, it also adds points about what makes a good document exchange format. It also discusses the values of standards of exchange in terms of establishing a level playing field. The Galician is pretty much readable to those who can read Spanish.

<http://members.tripod.com.br/ramonflores/word/index.html>

**Brave new Word** A similar document in Norwegian, a language I can't read. Written by Thomas Gramstad. It has some links at the end that might be useful to people who don't read Norwegian.

<http://www.efn.no/brave-new-word.html>

**Avoid E-Mail attachments, especially Microsoft Word** A similar document to this, but much shorter. It gives some brief instructions to MS-Word users on alternatives they can use for document exchange.

<http://bcn.boulder.co.us/~neal/attachments.html>

## **Elektronische infomatieoverdracht binnen de VU-organisatie: Het gebruik van e-mail**

A document in Dutch by Reinout van Schouwen. Also it is directed internally.

<http://www.cs.vu.nl/~reinout/word-attachments.pdf>

## **5.2 Rants about MS-Word**

The focus of this document has been on the misuse of Word for document exchange. It is geared toward MS-Word users to encourage them to send documents in other formats, even if they continue to use Word for document production. It should be noted, however, that those individuals who are most annoyed by receiving MS-Word files for document exchange are those who do not regularly use MS-Word. None the less, it is hoped that fans of MS-Word will recognize that whatever its virtues, it is not a document exchange format.

The arguments I've presented stand even if MS-Word were a good tool for document preparation. However, I'd also like to point to some documents which argue (correctly in my view) why MS-Word is a bad choice of document preparation system and not just a bad choice of document exchange format.

***Word Processors: Stupid and Inefficient*** by Allin Cottrell discusses what is wrong with What You See is All You Get systems using visual mark-up, as opposed to the far more reasonable structural system where you separate the tasks of controlling the appearance from the task of writing the content.

<http://www.ecn.wfu.edu/~cottrell/wp.html>

**No Proprietary Binary Data Formats** by Sam Steingold. This discusses the dangers of keeping important data in formats which require restricting and licensed software to recover. MS-Word is a proprietary and secret document format. You are trusting your future access to you own documents to the whim of a persistent monopolist.

<http://www.podval.org/~sds/data.html>

## **5.3 Reaction so far**

As far as I can tell my campaign has met with little success so far (January 2002) other than a few people taking some care to send me RTF documents

instead of MS-Word documents, with no change in their general practice. If I get any response at all it is typically “Well, you’re right but I’m going to stick with my current practices.” I find that disappointing, particularly when people acknowledge the correctness of the ethical argument I make.

On September 13, 2002 an opportunity fell into my lap during a discussion of a newly reported security bug in MS-Word to shamelessly plug this document in <http://slashdot.org/comments.pl?sid=39860&cid=4252157>. This generated a number of supportive email messages and a flurry of typo corrections.

There has also been one, somewhat harsh, critique of version 1.27 of this document. That critique and brief discussion can be found at <http://slashdot.org/comments.pl?sid=39860&cid=4264355>. I have modified the wording of section 1.9 and further emphasized the point made at the beginning of section 5.2 as a result.

## 5.4 How you can help

There are a number of ways you can help. These include, but are hardly limited to

1. Don’t use MS-Word for document exchange
2. Refer people who assume that you do use MS-Word for document exchange to this or similar document.
3. Promote the ideas described in this document. You may do this by linking to it or redistributing it. See section 5.5 for copyright notice and redistribution restrictions.

## 5.5 About this document and copyright notice

This document is available in several formats from <http://www.goldmark.org/netrants/no-word/>.

Copyright (c) 2001–2002 by Jeffrey Goldberg. This material may be distributed only subject to the terms and conditions set forth in the Open Publication License, v1.0 or later (the latest version is presently available at <http://www.opencontent.org/openpub/>). Distribution of the work or derivative of the work in any standard (paper) book form is prohibited unless prior permission is obtained from the copyright holder.

Please note that that if you wish do something with this that requires my explicit permission, just ask. I suspect that I'd grant it for most requests. Note also that the Open Publication License does allow you to do many things with this document without my permission.

## **5.6 Shameless plug**

If you have found this interesting, you may wish to see other netrants I have at <http://www.goldmark.org/netrants/>.

## **5.7 Acknowledgement**

Among others, I would like to thank Jim Diamond, Alan Frame, Dave Reader, Pete Mitchell and Juhapekka Tolvanen for their comments on an earlier draft. Your name can be added here as well. Just provide useful comments and suggestions. Other people are acknowledge in the change log of this document.