

The world does not need a "conversion nightmare": a standard office file format already exists



Published on Free Software Magazine (<http://www.freesoftwaremagazine.com>)

The world does not need a "conversion nightmare": a standard office file format already exists

Why the world needs to say no to OOXML, and should use ODF instead

By Tony Mobily

This is an editorial about file conversions. It starts with a story about Free Software Magazine and our struggle with article formats, and continues explaining why the world needs to get rid of Office Open XML, which could create more problems than the Microsoft monopoly itself.

When I started Free Software Magazine, we faced the problem every publication needs to face: which file format should we use for articles? It was a few years ago now (as they say, time flies when you have fun!). At the time, the web site wasn't our main focus: we were actually printing a paper magazine (!), we were generating amazing PDF files using LaTeX, and decided that a static web site was going to "do" for quite a while. We decided that the "master" format for our articles would be XML. XML seemed like a good idea at the time. None of the other options seemed quite as feasible: text wasn't enough, HTML was too vague, ODF was too complex, and so on. Plus, everybody was using it.

Since we couldn't find a single decent semi-visual XML editor, we asked our authors to hand in XML directly. Of course, people became very creative when they created an article file: we had to write a script that deleted white spaces around tags, and generally "cleaned up" the XML files we received. We also had to check manually that the files had the right em dashes, the right opening and closing speech marks, the right apostrophes, and so on. I won't even get started on the problems some authors had with getting the XML right: <p> tags left unclosed, items without first, and so on. It doesn't sound complicated, but when you have a 2500 word article full of listings, text boxes, figures and so on, and (even worse) when the XML error you get from the parser is as unhelpful as it could be, things got tricky. It was a small nightmare, which repeated with every issue of the magazine, and nearly every article. Two prospective (and influential) bloggers refused accounts with Free Software Magazine when they realised they would have to spend time tagging up XML files. Laziness? Maybe. But, as we say around here, "fair enough".

Luckily, the delirium is now over. We have **upgraded** our article format to Markdown Extra (although, it has a few tweaks to allow tables and textboxes). Authors can now write articles following this Free Software Magazine article template. Issue 21, this very issue, was edited mainly using the new file format.

Converting the articles from XML to Markdown Extra/FSM was a lot of hard work. I just about managed to do using XSLT with custom PHP calls within the XLS file. (If you are thinking "the XSLT from a basic format to Markdown should be simple", I will give you a few keywords: "white paces", "enters", "tables", "clashing escape characters", "CDATA", and so on). The conversion required substantial trial-and-error and tweaking. It contains several hacks I am not especially proud of. To date, I am not yet 100% sure it actually works for every single article. And we are talking about translating an *extremely* simple XML format into an

The world does not need a "conversion nightmare": a standard office file format already exists *extremely* simple text format. As always, the conversion part was easy. However, getting it to actually work was tricky.

This change won't affect you —well, apart from the occasional due to the occasional hard-to-translate article (we have over 2000 articles in our database, and we checked things by "statistical sampling"...). What is interesting is that this adventure (which I named "article conversion hell") reminded me of something that sounds obvious, but we tend to forget: file conversions are complicated, sub-optimal, time-consuming, imperfect by nature, often wrong, often the result of guess-work, tricky, and basically evil. When you open a Microsoft Office 2000 file using OpenOffice, things might work seamlessly, things might look a little odd, the file might look perfect—but if saved back as a Microsoft Office 2000 file, it might be ruined forever. There is a reason for this: file conversions need to be avoided (especially, like in this case, if the original file is an undocumented back-back-back-back-backward compatible format which really doesn't deserve to exist anymore, and didn't deserve to exist in the first place). ODF isn't perfect (yet?), but it aims at being the format for office documents. It's standard, and several pieces of software today can handle it (see: it's not an OpenOffice-only game).

Microsoft trying to shove OOXML down ISO's throat (and effectively damaging, maybe beyond repair, the image of what should be an independent body) can damage the computer industry immensely. The fact that both ODF and Office Open XML are XML means absolutely nothing. You can see here a [technical comparison between the two](#): converting one format to the other is anything *but* fun. Thousands of bogus documentation pages that come with OOXML don't help.

What I experienced with Free Software Magazine while converting (which, admittedly, wasn't really that big a deal) would be nothing compared to what the whole world will have to deal with if OOXML became "the" file format "normally" used to exchange office documents. A situation like this will impose constant conversions, quirks, compatibility problems, and so on all of us. It will also be a fantastic card for Microsoft: "look, GNU/Linux is sort of good, but you know, you can never trust it to open an XML file... sometimes the images are squint, you know..."

Microsoft knows this. Unsurprisingly, they have recently announced that they would release several conversion tools to translate ODF into OOXML and vice-versa. I read the article right in the middle of my "article conversion hell", and wondered if anybody else realised how disastrous it would be, if Microsoft managed to convince the world that it was "OK" to have two competing standards, since it's so easy to convert them into each other. The risk is very real: if we don't stop them, Microsoft will muscle its way in, and will force the whole world to fight with conversions for years, or decades, to come.

Microsoft proposed a bogus Office file format while an ISO standard already existed. Their shady practices to get their format fast-tracked and approved by ISO didn't work. But Microsoft is still trying—and I can guarantee, it will keep on trying until it succeeds.

The only possible answer for Microsoft and OOXML is simple: the world already has an office file format. The world doesn't need nor want a "conversion nightmare". The world's ISO-approved Office format already exists: it's called ODF. Microsoft: deal with it!.

Biography

[Tony Mobily](#) (/user/2" title="View user profile.): Tony is the founder and the Editor In Chief of Free Software Magazine

The world does not need a "conversion nightmare": a standard office file format already exists

Copyright information

Source URL:

http://www.freesoftwaremagazine.com/articles/editorial_22
