

Opening up Microsoft

Glyn Moody has posted [Open Science, Closed Source](#) in which he takes John Wilbanks to task for collaborating with Microsoft and effectively perpetuating Microsoft's stranglehold on word processing.

I agree with this analysis:

Working with Microsoft on open source plugins might seem innocent enough, but it's really just entrenching Microsoft's power yet further in the scientific community, weakening openness in general - which means, ultimately, undermining all the other excellent work of the Science Commons.

It would have been far better to work with OpenOffice.org to produce similar plugins, making the free office suite even more attractive, and thus giving scientists yet another reason to go truly open, with all the attendant benefits, rather than making do with a hobbled, faux-openness, as here.

Looking at the [example here](#) and [reading Pablo's Blog](#) I share Glyn Moody's concern. They show a chunk of custom XML which gets embedded in a word document. This custom XML is an insidious trick in my opinion as it makes documents non-interoperable. As soon as you use custom XML via Word 2007 you are guaranteeing that information will be lost when you share documents with OpenOffice.org users and potentially users of earlier versions of Word.

For something like an ontology this is completely unnecessary – all you should need to do is *link* to a web endpoint for a term to associate a word in your text with an ontology. This is similar to something I discussed with [Les Carr](#); a repository like ePrints could provide endpoint pages for links that when you link to them say “Les Carr is an author of the document that links to this”, that is, you bundle the predicate and the object together in an RDF assertion. All the users have to do is link to a page. That is a nanofORMAT that will work with *any* web capable tool on the planet, including wikis and text editors. Am I missing something here about how this could work at the file format level?

On top of the interoperability it might be really useful to have some custom code embedded in Word to help people *apply* these links, and if Microsoft are involved in that and the stuff is available open source then that's fine – it might help them sell a few more copies of Office, but on the basis of **ease of use** not **impossibility of escape**. Sun could even be involved in a similar effort for OpenOffice.org or Star Office if they wanted – they just don't seem to care too much about eResearch or [getting their word processor to talk to the web](#) at the moment.

[Another example I have been vocal about](#) is the [Microsoft NLM XML Add-in](#). If it works to allow ordinary word users to create XML to the NLM schema then it too will be one of these open-yet-closed Microsoft systems. Open source, yes, based on an open format, yes¹, in fact in this case two open formats yet it will only run in Word 2007 on the Windows platform and the source documents you create with it will only work on that platform. It may be useful but it will also continue the Microsoft stranglehold that Glyn Moody was complaining about. I just don't buy the argument that they can't implement the same thing using styles and have something that would at least interoperate with other word processors (including other Word versions).

I think we are seeing a new kind of format lock-in; a kind of monopolistic wolf in open-standards lambskin. I'm not saying that this is deliberate, at least not at Microsoft Research where the staff seem to be well meaning, open, communicative and friendly. They keep talking to me even though I keep ranting at them, at least so far.

1 Glyn Moody is adamant that the Office format is a psuedo standard that has harmed the ISO Standards forum. That may be so. Me, I welcome having the format well specified.

I warned Jim Downing from Cambridge about this kind of lock-in when I was over in Cambridge last year and he has taken up this issue with the Microsoft developers working on Chem4Word as [discussed here by Peter Murray-Rust](#), who also offers [this in defense of the work they're doing](#) there and [a follow-up](#).

In conclusion I offer this: I would consider getting our team working with Microsoft (actually I'm actively courting them as they are doing some good work in the eResearch space) but it would be on the basis that:

- The product (eg a document) of the code must be interoperable with open software. In our case this means Word must produce stuff that can be used in and round tripped with OpenOffice.org and with earlier versions, and Mac versions of Microsoft's products. (This is not as simple as it could be when we have to deal with stuff like Sun refusing to implement import and preservation for data stored in Word fields as used by applications like EndNote.)

The [NLM add-in](#) is an odd one here, as on one level it does qualify in that it spits out XML, but the intent is to create Word-only authoring so that rules it out – not that we have been asked to work on that project other than to comment, I am merely using it as an example.

- The code must be open source and as portable as possible. Of course if it is interface code it will only work with Microsoft's toll-access software but at least others can read the code and re-implement elsewhere. If it's not interface code then it must be written in a portable language and/or framework.