

DOC, PDF and XML - The Future of Documents as revealed through Acronyms

Date: 16th December 2005

Lawyers work with documents - increasingly in electronic form (as bytes) rather than in hard copy (as atoms). So, it is at the future of documents that I am looking for the next 12 months.

In the recent past, documents have broadly been in either Office (.doc) or Adobe (.pdf) format.

Each product was intended by its creators to provide a de facto standard for (and thereby a monopoly control of) the market. But each - originally - was intended to do different things.

Word was designed as a text input and manipulation medium. Acrobat was designed for text output and display. Or to put it another way: Word as a document creator and Acrobat as a document publisher.

Each has achieved dominance in its own area. We use Word to create documents and then to edit and collaborate in relation to revisions of them. Then, when a final version is agreed, we use Acrobat to publish it to others who cannot edit it but can only read it.

As well as differences in function, the commercial model for each is subtly different. Word is a straightforward 'everyone pays up front' product. Of course one can use the free Word Viewer but the whole purpose in using Word is to process text and not just to view it. So one has to buy it. Contrast Acrobat. Adobe's plaintive tagline 'There's more to Acrobat than the Viewer' says it all really. Here only a (relatively) few people pay up front to get the full product. For most people the purpose of an Acrobat document is to view it as a published unalterable item. So there is no need to buy it. Instead we use the free viewer. Worse, from Adobe's viewpoint, as the Acrobat format is openly published and available (though not open source) others can develop applications that eat into the market at which the full version of Acrobat is aimed.

Despite, or because of, the differences in origins and commercial model, gradually the two formats have begun to converge and conflict.

Primarily by Acrobat adding text manipulation features (including text recognition, editing, formatting, commenting and versioning) and, crucially, in later versions up to Version 7, interfacing with databases increasingly through XML. That left Word exposed to the risk of Adobe stealing its market dominance in document creation since, increasingly, Acrobat could do creation and publishing but Word could only do creation.

That risk is being addressed by Microsoft in a perhaps surprising way given historical views of Microsoft's strategy regarding intellectual property and file formats. At its heart, although openly published, Adobe's Acrobat was and remains a proprietary format controlled by Adobe. Until recently the same could be said of Microsoft Office in general and Word in particular.

But contrast Office 12.

Not only does this take XML as its native file format (with all the openness and interoperability that entails) but also Office 12 will output to PDF.

Potentially this is game set and match to Microsoft. Existing Office users will be able to create, manipulate and share Office (ie XML) files with:



- other people (office users or not - but why change from Office if it does what you want),
- other applications (so it will be as easy to open a word document in Internet explorer as in Word);
- databases (Word/XML documents will, much more easily than now, be able to take information from and put it into databases).

And then, when the document needs to be published in a fixed format - because it is final - they will be able to create a pdf version which, courtesy of Adobe's proprietary format is fixed. Result, everyone continues to buy Office (revenue for Microsoft) and to use the free Acrobat viewer (no revenue for Adobe).

For the inveterate Microsoft bashers this may seem a bad thing. For those with an eye to a return on investment next year it may not. Yes, there may need to be a gradual move towards Office 12 but the alternative would have been a move to Acrobat 7. And the result will be a familiar user environment but one delivering potentially massive gains in data sharing (via XML) and security (via PDF).

Will it happen? Time will tell.

