ALL IN THE XML

Dear Linux Magazine Reader,

Joe Casad, Editor in Chief

The State of Massachusetts has decided to require open file formats for office applications. The move, it seems, could have the effect of driving Microsoft Office out of official state business.

The state has thrown its support behind OpenDocument, an XML-based system that will appear in OpenOffice.org 2.0. It would not make sense to recount the whole discussion, but suffice it to say that many onlookers who don't usually comment on this kind of thing were genuinely confused. Even if you like open formats, why would you place such a large investment in a standard that just appeared and is only implemented in applications that are still in beta? But as always, the high-volume rhetoric drowned out the real issue.

Six years have passed since Microsoft announced that Internet Explorer 5 would be "the first commercially available browser software to support the Extensible Markup Language (XML) 1.0…" Microsoft adopted XML in a big way because, as they put it, "XML provides a universal language for data interchange..." And they didn't stop with tooling IE for XML. They went on to develop an entire XML-based Web service infrastructure and devoted innumerable conferences, white papers, tech tips, and marketing presentations to the premise that XML is a really good way to pass data from one application to another application.

At the same time, though, they were investing in another vision from the days when each application stored its data in a dark soup of numbers that no other application could fully interpret or manipulate in a competitive way. Microsoft's Office formats were the last great flowering of that vision.

Microsoft continued with these closed formats for business reasons even though they knew their days were numbered. The appearance of OpenDocument format may seem sudden or unexpected to the recently arrived, but the point is that the technology behind OpenDocument has been in the works for years. OpenDocument is simply a specification based on XML, which all the major infrastructure vendors (including Microsoft) fully support and celebrate.

Microsoft can no longer deny the limitations of hidden proprietary formats, and they plan to unveil new XML-based formats with the next version of Office. In other words, the age of secret proprietary formats is already over, and any discussion of the relative merits of OpenDocument versus traditional .doc and .xls files is totally irrelevant. Even Microsoft admits that the future is with XML, and they have moved to replace their binary .doc .xls formats with new XML-based equivalents called .docx and .xlsx. But they are unwilling to surrender control and have apparently placed license restrictions on the new formats that fail the Massachusetts definition of an open standard, thus crippling the formats as a "universal language for data interchange."

Microsoft's new formats are what you might call *encumbered XML*. The engineers write for a high level of interoperability, and the lawyers take it out again at the contract phase. This is yet another business move that Microsoft is entitled to make, but the buyer gets to do business too, and what government would ever agree to encumbered XML when it could have pure, ordinary XML with a license that is consistent with the promise of the technology?

The other point no one mentions is that governments create specifications all the time, and vendors complain about them all the time. They complain as long as it is in their interest, then when it looks like they aren't going to win, they suddenly discover the ability to adapt. In this case, maybe it is time for Microsoft to adapt, since it would be a trivial matter to support file formats that meet the Massachusetts guidelines. But then they would have to admit that grandiose names like ".docx" and ".xlsx" don't really mean that much anymore – it is all in the XML.

Joe