

---

Subject: Re: Small bug in DeXml

Posted by [phirox](#) on Sat, 07 Mar 2009 00:07:23 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Quote:The trouble is, that even if we recode LF as `&#x0A;`, you still aren't guaranteed you'll parse back exactly what you wrote into the XML previously, unless you use the `xml:space="preserve"` attribute. I believe that under normal circumstances you should use this attribute to ensure that the delimiting whitespaces are honored when parsing XML text elements.

In this case you are actually guaranteed that the LF character gets through since it is escaped. Also the `xml:space="preserve"` attribute only works as the name implies; for whitespace. All end of line characters should even be normalized to `#xA` according to the specifications, so there is no way to ever preserve any combination. See [here](#).

I can't think of any reason you wouldn't want to escape it. Not doing so actually breaks the current usability and is outside of the specifications.

Quote:Although the XML standard specifically doesn't forbid it, I personally find the concept of storing arbitrary binary data sequences as `&#-`escaped plain text into XML at the very least questionable; after all, we're trying not to invent our own XML but to use the one that's supposed to be a widespread standard, i.e. readable by non-U++ software without difficulties and possibly without semantic differences as well.

Even though this is quite unrelated to the bug; there is no other way to officially do binary encoding in XML than the `&#-`escapes. Doing something like a BASE64 function is actually "inventing your own XML" which btw is what XML is all about; customization.

---