

## EML - Bug #3465

### Change InlineType to CDATA

07/28/2008 12:10 PM - Jing Tao

|   |                    |                        |            |
|---|--------------------|------------------------|------------|
| <b>Status:</b>  | New                | <b>Start date:</b>     | 07/28/2008 |
| <b>Priority:</b>  | Normal             | <b>Due date:</b>       |            |
| <b>Assignee:</b>  | Matt Jones         | <b>% Done:</b>         | 0%         |
| <b>Category:</b>  | eml - general bugs | <b>Estimated time:</b> | 0.00 hour  |
| <b>Target version:</b>  | Unspecified        | <b>Spent time:</b>     | 0.00 hour  |
| <b>Bugzilla-Id:</b>   | 3465               |                        |            |
| <b>Description</b>  |                    |                        |            |
| Currently in eml definition, the InlineType, which is the data type of element inline, is PCDATA. This means xml parser will parse it. This causes some problems. For example, SAX parse in Metacat will change "<", if it is in element inline, to "<" during the parsing. This change will make the document invalid. If we change InlineType to CDATA, SAX parser will skip the content in element inline, the problem will be gone. |                    |                        |            |

### History

#### #2 - 08/20/2008 03:45 PM - Margaret O'Brien

This document

<http://dev.nceas.ucsb.edu/knb/metacat/mobrien.1.4/xml>

was inserted to metacat with the script wrapped in a cdata section, as in:

```
<inline>
<script>
<![CDATA[
function matchwo(a,b)    {
if (a < b && a < 0) then {
return 1;
}
else {
return 0;
}
}
]]>
</script>
</inline>
```

When metacat returns the doc, the CDATA wrapper is no longer there, the < etc get resolved and result is no longer valid xml. As expected, if I replaced the < and & with references AND wrapped the script in CDATA, metacat had no problem returning the whole doc, since now it has escaped characters to replace, too(see docid=mobrien.1.5).

So it seems that metacat is altering the incoming doc by removing the CDATA delimiters before it stores the content. Does this mean that this is not an EML problem, but a metacat behavioral problem instead?

#### #3 - 08/20/2008 04:08 PM - Margaret O'Brien

Sorry, that comment lost some text:

When metacat returned doc mobrien.1.4, the '<' etc that were originally enclosed in CDATA are now exposed, and the doc is no longer valid xml.

If instead of using a CDATA section, I replaced the offending characters with entity refs, (<) then predictably, these were resolved and the doc is no longer valid. See mobrien.1.6

Doing both (using entity refs inside a CDATA section, mobrien.1.5) preserved the document for viewing, but the returned xml doc is still altered by metacat - not a good thing.

#### #4 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3465

### Files

|              |         |            |                  |
|--------------|---------|------------|------------------|
| obrien.1.xml | 1.77 KB | 08/20/2008 | Margaret O'Brien |
|--------------|---------|------------|------------------|