

## Morpho - Bug #371

### problems with wizard-generated eml-physical documents

12/08/2001 01:18 AM - Matt Jones

<b>Status:</b>	Resolved	<b>Start date:</b>	12/08/2001
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Dan Higgins	<b>% Done:</b>	0%
<b>Category:</b>	morpho - general	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Beta2	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	371		

#### Description

The text import wizard generates eml-physical documents. They have some issues:

- 1) fieldDelimiter and recordDelimiter are written out in English instead of using actual ASCII symbol (e.g., "comma" instead of ","). Need to decide on a convention. I would suggest using the ascii character, or its octal (\054) or hex (0x2C) equivalents for non-printing characters. thus, a newline would be "0x0A" and a carriage return would be "0x0D", but a comma would be ",".
- 2) Wrong number of header lines entered into metadata. The file says there was 1 header line, when in fact there was 0.
- 3) The "size" element contains the string "bytes" which is not supposed to be there because it is defined in the "unit" attribute of the size element. So, it should be:  
<size unit="bytes">26</size>
- 4) the SYSTEMID in the DOCTYPE line is wrongly set to "eml-entity" when it should be set to "eml-physical"

#### History

##### #1 - 12/10/2001 10:21 AM - Dan Higgins

Concerning Item#1: The use of words to describe delimiters seemss appropriate (and will happen) as long as we allow the user to enter eml-physical information by hand. Only a computer geek will understand octal or hex (but such may be required for machine parsing!) - How do we specify a 'space' other than by the word or octal or hex?

##### #2 - 12/10/2001 10:53 AM - Dan Higgins

Concerning # of header lines::: Is the line containing column tiles (if present) considered a header line?

##### #3 - 12/10/2001 12:00 PM - Dan Higgins

items [#2](#), [#3](#), and [#4](#) in original comments have been fixed

##### #4 - 12/10/2001 12:36 PM - Matt Jones

Yes, the line containing column names is a header line, as are any other line that precede the actual data. In the case I described, there were no column headings (I unchecked the "use first row as headings" box), and so the number of header lines should have been 0.

As far as symbols for non-printing characters go -- I think the convention should be something that is universal, and therefore an octal or hex value. Its not pretty, but I think it is better than making up word symbols for all of the non-printing characters that exist in Unicode. User interfaces can always translate the non-printing characters into a symbol (like MS Word does for spaces, tabs, and hard returns, or as a word like "space"), but the actual XML should contain a more universal representation. Interestingly, XML provides a mechanism to use the hex value for Unicode characters, as does url-encoding for http, so maybe we should just go with that (hex). Its really just a matter of defining it in the EML documentation for eml-physical and then using that convention in Morpho.

##### #5 - 12/11/2001 05:13 PM - Dan Higgins

Field delimiter notation changed to XML hex notation; i.e. " " for a tab

**#6 - 03/27/2013 02:14 PM - Redmine Admin**

Original Bugzilla ID was 371