

EML - Bug #557

paragraph tag needs formatting structure

08/08/2002 08:10 AM - Matt Jones

| | | | |
|---|--------------------|------------------------|------------|
| Status: | Resolved | Start date: | 08/08/2002 |
| Priority: | Immediate | Due date: | |
| Assignee: | Matt Jones | % Done: | 0% |
| Category: | eml - general bugs | Estimated time: | 0.00 hour |
| Target version: | EML2.0.0rc1 | Spent time: | 0.00 hour |
| Bugzilla-Id: | 557 | | |
| Description | | | |
| <p>Bergsma suggested that the paragraph tag needs structure to accomodate real-life text such as lists and hierarchies. Chapal seconds the need. I tend to agree, but am not sure currently how to accomplish this. The viable proposals from my perspective seem to be:</p> <ol style="list-style-type: none">1) allow XHTML inside paragraph tags2) allow Docbook or simple Docbook in paragraph tags <p>Other proposals that seem less viable include:</p> <ol style="list-style-type: none">3) Decompose structured text into a series of <paragraph>.4) Inject structured text, with its native markup, as a CDATA block in <paragraph>.5) Make <paragraph> nestable <p>Three (3) and five (5) don't seem to solve the problem completely. Four (4) solves the problem but isn't at all standardized, and so interpreting paragraphs would be essentially impossible depending on what people used for their markup. Every author could use their own markup if we sanctioned (4).</p> <p>Here's a copy of the email thread that lead to this bug:</p> <p>---- Start message from scott.chapal at jonesctr org -----</p> <p>Has there been any discussion regarding Tim's questions? I thought he pointed out an area that deserves clarification before EML 2 is released.</p> <p>I too, am thinking there ought to be additional structure to represent prose.</p> <p>Tim Bergsma <tbergsma@kbs.msu.edu> writes:</p> <p>The problem, as I have mentioned previously, is that prose metadata (text) is often highly structured. <paragraph> gives us no way of representing the structure of text, which is itself information. In many instances, of course, <paragraph> is repeatable, which allows us some leeway to represent sequential structure. But there is still no way to represent hierarchical structure. This has significant consequences. For example, a project-level abstract may include a short outline of purposes or hypotheses. A research protocol may include finely-grained outlines of contingencies and responses.</p> <p>Three alternative solutions have emerged from previous discussion.</p> <ol style="list-style-type: none">1. Decompose structured text into a series of <paragraph>.2. Inject structured text, with its native markup, as a CDATA block in <paragraph>.3. Make <paragraph> nestable. <p>...</p> | | | |

I hope that the leadership of the eml development community will offer me some guidance on this issue. I really don't think number 1 is a viable option, but could make peace with either 2 or 3.

I don't particularly like any of those options.

How about deferring to some [SGML/XML] standard to represent prose?

Possibly DocBook; use the (W3C)Schema when it is complete?

Or even the simplified DocBook?

<http://www.oasis-open.org/committees/docbook/specs/wd-docbook-simple-1.0-CR2.html>

Other??

Necessarily, existant documentation would have to be deconstructed or converted to a markup language format. Or if visual formatting is paramount then it could point to a (quasi) neutral file format like .pdf, but that wouldn't accomplish textual indexing, querying and structure that EML aims to support.

This comment of Tim's really struck me:

However, converting hierarchically-structured text to serially-structured text will require innovations by the data manager that raise him/her to the status of author, a status not necessarily sanctioned by those who contributed the original material.

I think this concern is largely unwarranted. Were it viable, Information Scientists, 'New' librarians and editors wouldn't be able perform their functions either. Fidelity to an original work should be mandatory, but the translation process should be made transparent and trivial. If an editing review is needed, then create one. Perhaps even a 'stamp of approval' or something. We're talking about metadata -- **documentation**, after all. If it's not structured it's really not very useful.

---- End message from scott.chapal at jonesctr org -----

Related issues:

Has duplicate EML - Bug #558: paragraph tag needs formatting structure

Resolved

08/08/2002

History**#1 - 08/26/2002 01:50 PM - Matt Jones**

Resolving as duplicate, looks like I entered it twice accidentally.

- This bug has been marked as a duplicate of 558 ***

#2 - 03/27/2013 02:14 PM - Redmine Admin

Original Bugzilla ID was 557