

EML - Bug #3181

xs:string to ComplexType TextType, minOccurs=0, judiciously applied

03/21/2008 04:13 PM - Margaret O'Brien

Status:	New	Start date:	03/21/2008
Priority:	Normal	Due date:	
Assignee:	James Brunt	% Done:	0%
Category:	eml - general bugs	Estimated time:	0.00 hour
Target version:	Postpone	Spent time:	0.00 hour
Bugzilla-Id:	3181		

Description

This is a summary of a recent discussion on eml-dev which does not appear to have been entered in bugzilla. Several people have expressed a need for additional structure in leaf nodes that are currently designated xs:string, generally to accommodate formatting for species binomials, chemical notation and lists. Examples include <title>, <method>, and <protocol>.

One solution is to change these from xs:string to txt:TextType. Since TextType is mixed content, it will not affect existing documents containing strings. The nodes to apply this change should be agreed on by this group, and this is not meant to be a work-around for eml which needs enhancement. Database implementations will need to correctly interpret the data typing when searching these elements. For more info on TextType, see bug 2703, and the docbook schema (<http://www.docbook.org/specs/>).

EML 2.0.1 title element:

```
<xs:element name="title" type="xs:string" maxOccurs="unbounded">
```

EML 2.0.2 proposed title element:

```
<xs:element name="title" type="txt:TextType" maxOccurs="unbounded">
```

Either of these is valid:

```
<eml>
<dataset>
<title>Uptake of nitrogen by Alnus tenuifolia and Alnus crispa in six different successional habitats</title>
...
</dataset>
</eml>
```

```
<eml>
<dataset>
<title>Uptake of nitrogen by
<emphasis>Alnus tenuifolia</emphasis> and
<emphasis>Alnus crispa</emphasis>
in six different successional habitats</title>
...
</dataset>
</eml>
```

History

#1 - 03/21/2008 04:29 PM - Margaret O'Brien

changed summary only, -- this is not a change to txt:TextType, since that would require adding children elements (para or section) late friday.

#2 - 09/22/2008 12:20 PM - Margaret O'Brien

targeting for 2.1.0, although may drop back to unspecified.

#3 - 10/01/2008 09:57 AM - Margaret O'Brien

Summary 2008-09-30:

Matt Jones, Margaret O'Brien, James Brunt, Mark Servilla, Inigo San Gil, Chris Jones, Corinna Gries, Ken Ramsey

Retyping from xs:string to TextType to allow formatting in certain fields is not an ideal solution for 2 reasons (also discussed in previous comments to this bug)

1. any embedded elements within titles and abstracts will create barriers to searching in both knb and other catalogs, and
2. the practice of adding visual formatting to EML does not conform to the basic principle that XML should encode semantics not style.

A better solution might be to identify the terms which need additional formatting and create structures to accommodate them which can be transformed as needed.

Known terms include species binomials and chemical formulae, which might be surrounded by tags such as <speciesBinomial> and <chemicalFormula>. <speciesBinomial> would have simple content that could be transformed (e.g., to <emphasis> in stylesheets). Authors would need to employ another schema (e.g. ChemicalML) inside <chemicalFormula>. Other types of terms used by EML authors will need to be identified.

A subgroup of eml-dev (James Brunt, Chris Jones) will consider possible solutions. Current uses of txt:TextType in instance docs should be examined to determine whether these are semantic or presentational, and to identify terms to consider for additional structures. Other schema languages could also be examined (FDGC has no provisions for formatting resource titles).

#4 - 03/27/2013 02:22 PM - Redmine Admin

Original Bugzilla ID was 3181