

## EML - Bug #3445

### stmml.xsd non-deterministic

07/10/2008 04:37 PM - Margaret O'Brien

<b>Status:</b>	Resolved	<b>Start date:</b>	07/10/2008
<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>	EML2.1.0	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	3445		

#### Description

A section of stmml.xsd is invalid, according to a new parser feature that Jing just added in response to bug 3232, and also to the venerable parsers in xmlSpy and oxygen. Interestingly, the 2 commercial editors don't catch up the error unless the schema is loaded directly, instead of imported (ie, by attribute.xsd).

This bug is very similar to 2054 -- that plagued EML for so long.

Here is the offending snip from stmml.xsd, starting at line 1708. The problem is the unbounded "definition" right next to the <xs:choice>

It appears that the invalidity can be fixed by either removing the minOccurs (ie, making it required) or by removing the <element ref="definition" ...> altogether. The sequence followed by choice structure makes any combination of elements ok, so this declaration seems to be extra. I do not see a more recent stmml.xsd available (cml.sourceforge.org). And we shouldn't go trekking of with our own flavor of stmml, so awaiting recommendations.

```
<!-- sequence -->
<!-- choice -->
<!-- definition -->
<!-- alternative -->
<!-- annotation -->
<!-- description -->
<!-- enumeration -->
<!-- relatedEntry -->
</choice>
</sequence>
```

#### Related issues:

Is duplicate of EML - Bug #3448: stmml.xsd non-deterministic

Resolved

07/11/2008

#### History

##### #1 - 07/14/2008 11:41 AM - Margaret O'Brien

- Bug 3448 has been marked as a duplicate of this bug. \*\*\*

##### #2 - 07/14/2008 11:44 AM - Margaret O'Brien

I agree with Inigo, that in the long term, we need a better structure than the standard/custom unit situation. But in the short term, we need to easily migrate 201 docs to a valid schema.

One simple short term solution for stmml.xsd is to remove the offending element declaration. This wont affect 201 docs, since - as currently - they will be validated with the simple parser.

I think (but have not tested) that it wont affect 2.0.1 docs that are migrated to 2.1.0 either, because the sequence->choice structure allows these elements in any number, any order, and that first element declaration for <description> is superfluous.

Another option is to not use the schema-full-checking behavior in the 2.1 parser, which lets non-deterministic bugs slide by (see bug 3232).

But like I said, no potential fixes to stmml have been tested, and I cant get to it for a week or more, although someone else might. And first, we should contact the authors for their recommendation. My cursory search did not turn up development group for stmml, only contact info for the Murray-Rust group at Cambridge (and they seem to be deep into CML now).

**#3 - 08/08/2008 02:37 PM - Margaret O'Brien**

Regarding the conflicting <definition> elements:

After looking over the stmml documentation a little more, I think that the best solution for creating a valid stmml is to remove (comment-out) the <xsd:element ref="definition"/> inside the choice group, and leave the first declaration, <xsd:element ref="definition" minOccurs="0"/>.

I found a statement suggesting that <definition> is supposed to be "An almost mandatory child element of entry, giving a formal definition of the term" (line 1547), which could imply that the authors intended the "definition" to appear first, not just anywhere. This bit of documentation is not near the element declaration itself, but in a summary section for the parent element. However, I cant quite figure out how they intended to specify the conditions under which it was required.

I have asked P. M-L, but havent heard back. I will be gone next week and want to get this tested and checked in so others can use it in their work on metacat and the parser.

**#4 - 08/08/2008 04:19 PM - Margaret O'Brien**

Regarding the change proposed to stmml in comment [#3](#):

this change is not backward compatible with stmml1.0 docs. Removing <description> from the choice group, means that it is only allowed as the firstChild of entry, whereas before it could be placed anywhere.

Keep in mind that to the best of our (well, my) knowledge, EML doesn't use this part of STMML at all. This change affects the dictionary/entry tree (dictionary is another global element). We import the whole schema, but use only the unitList (also global). However, someone out there could be putting STMML dictionary trees into EML2.0.1, and those instance docs (if any exist) would not be valid EML2.1.0. This will be added to the release notes.

Regarding the namespace we use for stmml: Per a discussion on irc with Matt, we'll use a xmlns="http://www.xml-cml.org/schema/stmml-1.1" to be consistent with our policy of naming in EML.

I've written to the STMML's devs for their recommendation.

**#5 - 09/22/2008 10:10 AM - Margaret O'Brien**

changing status to "fixed"

**#6 - 03/27/2013 02:23 PM - Redmine Admin**

Original Bugzilla ID was 3445