EML - Bug #2076

containers for eml-literature docs

05/03/2005 11:49 AM - Margaret O'Brien

Status:	New		**Start date:**	05/03/2005
Priority:	Normal		**Due date:**	
Assignee:	Matt Jones		**% Done:**	0%
Category:	eml - general bugs		**Estimated time:**	0.00 hour
Target version:	Unspecified		**Spent time:**	0.00 hour
Bugzilla-Id:	2076			

Description

SBC-LTER has been investigating using EML and Metacat for our site's bibliography. In creating and displaying my xml, I've come up with a list of potential changes to the eml-literature schema, which dont affect the validation of current citation eml docs. At the network office, James Brunt and Mark Servilla are also pursuing something similar while constructing the network's litdb, and will have some input soon.

My goal was to create a hanging-paragraph style display of our site's bibliography, with links to the paper and to the dataset as appropriate. EML's imports/includes were a bit intimidating, so instead, I chose to start from scratch and write a dtd that looks like eml-literature, but included just the basic tags needed for the typical hanging-paragraph bibliography display. So the dtd is kind of an "eml-lite". But in the long run of course, my wimpy dtd should disappear.

What I've done so far: Create a test xml doc of several pubs with this dtd. The citations are mostly fictional to suit my stylesheet testing needs. I created a stylesheet to display it in a typical hanging-paragraph list, filtered by type, then sorted by year and author. Looking for some thrills, I went ahead and inserted it into metacat, and mapped the stylesheet (with help from Sid). This seemed to be a good way to show what I had in mind (not to mention, see if it really worked). Since most of the citations were created for testing the stylesheet, most of the url links really dont go anywhere. It was the general format that I was interested in. You can see the results at:
http://sbc.lternet.edu/catalog/metacat?action=read&qformat=sbclter&docid=sbc_pubs_test.1.1
The xml file and stylesheet can be found at:
http://sbc.lternet.edu/external/EML/SBC_publications/sbcCitationList.xml
http://sbc.lternet.edu/external/EML/SBC_publications/sbcCitationList.xsl

What I haven't done: some cleanup, create css parameters, use id refs, searches, link the title to the existing table-style eml-literature.xsl. I've started on a script to convert SBC's own html list of pubs to xml.

Here is the list of differences between my dtd and the eml-literature schema:
1. this dtd has a <citationList> as the root element, (eml: no such tag)
2. <citation> is child of citationList, 1 to many allowed. (eml: one only, root element)
OK, these 2 do affect current docs, unless they are put into another module (eml-publications?) that uses eml-literature. Having all the citations in one list is much easier to maintain than the current scheme of 1-citation-per-doc. Also, a research site is likely to have many repeated authors, and if the pubs are in one list, authors can be maintained in the additionalMetadata, and linked with ids.

3. <title> is allowed to have children, mainly so species binomials can be italicized. Changing <title> from type:string to type:text would take care of this (?without affecting current string content?)

4. journal, volume, pageRange are 0 or 1, to accomodate in_press/submitted papers. <pubDate> is already optional. My stylesheet used the absence of a pubDate to filter out the in-press pubs. Citations may spend only a short time

in this state, but it's very important to scientists to make their newest papers available quickly.

5. added an optional <contact> tree, since the first author is not always the person to contact for reprints. The stylesheet looks here first, then at the list of creators.

6. added an optional <datasetId> so an accompanying archived data package can be recorded. This is debatable. I was looking for a way to link archived datasets to the citation, since some journals are requesting that data be published along with papers. It seemed a better idea to start from the citation and link back to the dataset(s), rather than including the finished paper's citation with the dataset metadata, since after a dataset is revised, the paper may belong only with an earlier version. Also, papers are likely to use data from multiple datasets. I was partial to the datasetId tag because the rest of the url could be created with stylesheet variables. However, this method doesnt allow urls for any other data catalogs to be included (unless you made additional variables). Chris suggested an alternative - to use a <distribution type="information"> tree for the dataset. I'm not sure that this is specific enough.

7. added an optional <description> to distribution/online. Actually, I've wished for this (or something like it) in eml-dataset, too. It provides a place to put some text which can appear inside the anchor tags in the html. Which would really help if the dataset link was put here, to avoid having to diplay the ugly url and instead describe where the link actually goes. But maybe there's already a mechanism for this that I've missed.

Thanks-
Margaret (sbc-lter IM)

Related issues:

Blocked by EML - Bug #3163: eml-literature schema - cardinality of volume and...	**Resolved**	**02/26/2008**
Blocked by EML - Bug #3164: sbc-lter adaptation of eml-literature schema - co...	**Resolved**	**02/26/2008**

History

#1 - 09/22/2008 05:09 PM - Margaret O'Brien

This bug was split into individual bugs, since they will be addressed in different releases. This is the original report, and launched a discussion on collections of eml documents in general:
http://mercury.nceas.ucsb.edu/ecoinformatics/pipermail/eml-dev/2005-August/001124.html
http://mercury.nceas.ucsb.edu/ecoinformatics/pipermail/eml-dev/2005-August/001128.html
http://mercury.nceas.ucsb.edu/ecoinformatics/pipermail/eml-dev/2005-August/001129.html

#2 - 03/27/2013 02:19 PM - Redmine Admin

Original Bugzilla ID was 2076