

Infrastructure - Task #8753

Add support for EML 2.2 (indexing, view)

2018-12-19 00:43 - Bryce Mecum

<b>Status:</b>	New	<b>Start date:</b>	2018-12-19
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Bryce Mecum	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Story Points:</b>	
<b>Milestone:</b>	None		
<b>Product Version:</b>	*		
<b>Description</b> <p>EML 2.2 is nearly released and, with it, comes changes to indexing and the view service. The key points are that EML now supports Markdown and semantic annotations but some more stuff was added too. See the in-progress What's New in EML 2.2.0 page <a href="https://github.com/NCEAS/eml/blob/BRANCH_EML_2_2/docs/eml-220info.md">https://github.com/NCEAS/eml/blob/BRANCH_EML_2_2/docs/eml-220info.md</a> for details.</p> <p>I'll track these with sub-tasks but here's an overview:</p> <b>Formats</b> <p>We need to expand the list of formats to include EML 2.2.</p> <b>Indexing</b> <p>EML 2.2 requires adding new fields and changes to some existing fields. For example, abstracts in EML can now also include a markdown child element which we'll also want to store in Solr. We also want to add in support for semantic annotations and probably structured funding information since that's been requested so much.</p> <p>Support for semantic annotations will likely use the same approach as our previous work on semantic search where incoming annotations were subject to materialization during the indexing process. To explain what that means, the idea is that, when an annotation for term X comes in, the index processor loads the relevant ontology, materializes the superclass hierarchy for the term, and stores the entire hierarchy in Solr. For EML records with multiple annotations, only the unique set needs to be stored.</p> <b>View service</b> <p>As mentioned above, EML 2.2 changes how some existing elements (e.g., abstract) work and adds some new ones (e.g., annotations). The existing EML 2 stylesheets will work for EML 2.2 because 2.2 is backwards compatible but we'll want to extend them to support what's new and changed in EML 2.2</p> <p>Of note: EML now supports storing Markdown where previously only EML TextType (basically DocBook) was allowed. Our plan on Metacat/MetacatUI to support this is to render the Markdown on the client side which does involve some security concerns (see <a href="https://github.com/NCEAS/metacatui/issues/860">https://github.com/NCEAS/metacatui/issues/860</a> for tracking).</p>			
<b>Subtasks:</b>			
Task # 8754: Add EML 2.2 to CN formats list			New
Task # 8755: Expand EML indexing support for EML 2.2			New

History

#1 - 2018-12-19 00:44 - Bryce Mecum

- Assignee set to Bryce Mecum