

Infrastructure - Story #6848

Expand ObjectFormat.formatType options

2015-02-11 06:21 - Ben Leinfelder

Status:	Closed	Start date:	2015-02-11
Priority:	Normal	Due date:	
Assignee:	Ben Leinfelder	% Done:	100%
Category:	d1_indexer	Estimated time:	0.00 hour
Target version:	CCI-2.0.0		
Story Points:			
Description <p>I'm trying to support a semantic annotation object format - in this case it is serialized as JSON - and am unsure how to type it. In most ways it should be treated like a RESOURCE formatType in that:</p> <p>it should be harvested and stored on the CN</p> <p>it should be indexed by the CN</p> <p>indexing these objects affects index values for other objects that are described by the annotation model</p> <p>Unfortunately, we require (or assume?) that a RESOURCE formatType is an ORE model serialized as RDF/XML.</p> <p>We should relax this requirement using one or more of these options:</p> <p>allow other ObjectFormats to be be classified as RESOURCES and synchronize them without an extra "validation" step</p> <p>consider indexing DATA objects and type the annotation as that</p> <p>consider allowing non-XML METADATA objects to be stored and indexed and type the annotation as this</p> <p>consider adding a new formatType that is handled similarly to our current RESOURCE formatType but is more lenient and will allow for JSON serializations (and others)</p>			
Related issues:			
Related to Infrastructure - Task #3972: Validate DataONE Resource Map Objects		Closed	2013-09-11

History

- #1 - 2015-02-11 06:21 - Ben Leinfelder
- Related to Task #3972: Validate DataONE Resource Map Objects added
- #2 - 2015-06-16 15:13 - Dave Vieglaiss
- Target version changed from CCI-2.0.0 to CCI-2.0-TRUNK
- #3 - 2015-06-25 22:22 - Ben Leinfelder
- Assignee set to Ben Leinfelder
- % Done changed from 0 to 100

- *Category set to d1_indexer*
- *Milestone changed from None to CCI-2.0*
- *Status changed from New to Closed*

I've since refactored the indexing process such that we can handle RESOURCE format types as XML or as any other format - provided there is an appropriate subprocessor that can handle the format. In my case, the subprocessor parses the JSON format and looks for annotation tags. The existing ORE processo and the generic RDF/XML subprocessor continue to parse their formats as they wish.

#4 - 2015-07-01 16:22 - Dave Vieglaais

- *Target version changed from CCI-2.0-TRUNK to CCI-2.0.0*