

## CN Index - Feature #6719

### Update data package relationship processing wrt maintaining SID based relationships.

2015-01-06 17:19 - Skye Roseboom

<b>Status:</b>	Closed	<b>Start date:</b>	2015-01-06
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Skye Roseboom	<b>% Done:</b>	100%
<b>Category:</b>	d1_cn_index_processor	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	CCI-2.0.0		
<b>Story Points:</b>			
<b>Description</b>			
SeriesId use in data packages:			
1.) A ORE/data package defines relationships (documents, documentedBy) using seriesId to refer to data and science metadata documents.			
Should data package relationships copied to each document in a series or just on the current end of the series?			
After an ORE/data package/ which defines relationships using seriesID is indexed, a new version of a science metadata document in the series used in an ORE/data package - the seriesId based relationships need to be placed on the new document in the series. The series based relationships are removed from the previous documents in the series?			
Same question/issues for annotations or provanance info? Can annotations be seriesId based, and if so are they placed on each document in the series?			
Possible strategies:			
Model 'aggregates' relationship on resource map records in search index.			
Re-process all resource maps found on previous 'head of sid' chain when adding indexing document with sid value.			
Issues:			
Additions to SID chain will modify definition of data package after the ORE/data package document has been processed for index.			
Need to maintain current state of data packages.			

#### History

##### #1 - 2015-02-03 18:43 - Skye Roseboom

- Parent task deleted (#6716)

##### #2 - 2015-02-03 18:43 - Skye Roseboom

- Tracker changed from Task to Feature

- translation missing: en.field\_release set to 2

##### #3 - 2015-06-30 15:28 - Skye Roseboom

- Target version set to CCI-2.0-RC1

##### #4 - 2015-07-01 16:20 - Dave Vieglaais

- Target version changed from CCI-2.0-RC1 to CCI-2.0.0

##### #5 - 2015-07-01 18:00 - Dave Vieglaais

- Category set to d1\_cn\_index\_processor

##### #6 - 2015-07-16 19:48 - Skye Roseboom

- Description updated

**#7 - 2015-07-21 19:03 - Ben Leinfelder**

- *Status changed from New to Closed*
- *% Done changed from 0 to 100*

Now indexing OREs using both PIDs and SIDs. When objects (referenced by SID) in an ORE are updated, we submit the ORE for index processing again such that the package/documentation relationships are maintained for each revision of the object[s] in the sid chain[s]. See SolrIndexReprocessTest for details on our expectations.

**#8 - 2015-07-23 22:17 - Ben Leinfelder**

- *% Done changed from 100 to 30*
- *Status changed from Closed to In Progress*

Reopening as we rework the processing logic to only index the latest version of datapackge contents that use SIDs.

**#9 - 2015-08-05 20:02 - Ben Leinfelder**

Current behavior, as tested in unit test:

- \* Datapackage contains a mix of PID- and SID-identified objects.
- \* SciMeta is Identified with SID, then it is updated. Index then reflects that only the head-revision is in the datapackage and only the head revision "documents" the data objects in the package
- \* Data is identified with a SID, then it is update. Index then reflects that only the head-revision is in the datapackage and only the head revision shows "isDocumentedBy" pointing to the SciMeta identifier.

**#10 - 2015-09-22 16:18 - Skye Roseboom**

- *Status changed from In Progress to Closed*
- *% Done changed from 30 to 100*

Ben finished work on this and updated unit tests. Closing issue.