Infrastructure - Bug #8850

v2.3.11 RdfXmlSubprocessor / HttpService.getDocumentBySeriesId can be wrong

2019-11-06 18:29 - Rob Nahf

Status:	New	**Start date:**	2019-11-06
Priority:	Normal	**Due date:**	
Assignee:	Jing Tao	**% Done:**	0%
Category:	d1_indexer	**Estimated time:**	0.00 hour
Target version:	CCI-2.3.11		
Milestone:	None	**Story Points:**	
Product Version:	*		

Description

branch D1_CN_INDEX_PROCESSOR_v2.3 introduced a new method HttpService.getDocumentBySeriesId, to be used by RdfXmlSubprocessor to get the head of a series.

It is buggy in that it determines the head of the series based only on the value of the obsoletedBy field, and in the index, more than one solr document can fit that criteria. The simple case (and that is likely to be enriched in multi-threaded indexing) is when the systemMetadata update of an update of an object is processed at some future time after the indexing of the update itself. When indexing the update, both the update and the original would be returned from Solr, and the client method would return which ever one happened to be first in the list.

The v2.4 indexer which uses "total relationship enumeration," will not use this method, so fixing this bug is only important if the 2.3 logic will be maintained.

If it will be maintained, it's important to note that the DataONE spec does not require the obsoletedBy field to be populated, although Metacat does. Mutable MemberNodes my never update the obsoletedBy field of those objects not at the head. Similar logic implemented for resolve should be considered for these cases.

History

#1 - 2019-11-06 18:32 - Rob Nahf

- Description updated