

Infrastructure - Story #8504

Support creation of data citation record from solr record

2018-03-19 21:53 - Dave Vieglais

Status:	New	Start date:	2018-03-19
Priority:	Normal	Due date:	
Assignee:	Monica Ihli	% Done:	0%
Category:	d1_indexer	Estimated time:	0.00 hour
Target version:	CCI-2.4.0		
Story Points:	20		
Description			
The goal of this story is to ensure that elements in the solr search schema are available and appropriately populated to support generation of DataCite version 4.x or later records.			
By ensuring support for this schema, it can also be asserted that suitable citation metadata can be provided in landing pages and other renderings of content provided by DataONE.			
Resources:			
<ul style="list-style-type: none">DataCite Schema version 4DataONE solr Search fieldsRDA Data Citation Recommendations			
Subtasks:			
Task # 8505: Review solr terms for coverage of data citation elements			New
Task # 8506: Identify ISO metadata terms appropriate for population of data citation el...			New
Task # 8507: Support generation of citation metadata records from dublin core			New
Related issues:			
Blocks Infrastructure - Decision #8189: Proposal to change the roles mapped t...		New	2017-10-02

History

#1 - 2018-03-19 21:59 - Dave Vieglais

- Blocks Decision #8189: Proposal to change the roles mapped to the origin Solr field for ISO docs added

#2 - 2018-03-19 22:29 - Monica Ihli

Procedural notes:

1. Evaluate DataCite v.4+ requirements.
2. Identify currently indexed values for each formatID.
3. Identify and address any gaps between DataCite requirements and currently indexed elements.
4. Ensure that the citations in search.dataone.org display appropriately.

Also:

- Start with EML & ISO

#3 - 2018-07-17 22:41 - Bryce Mecum

Just a note, Chris Turner has been waiting on progress on this since August 24, 2017. When/if any progress gets made w/r/t ISO citations, it'd be good to email him to let him know what's changed and how it affects the RW node.

#4 - 2018-07-17 22:59 - Bryce Mecum

I re-read what I just wrote above and want to add two things:

- I was initially assigned and didn't get this done so tardiness is largely my fault
- Chris Turner is specifically looking for how ISO* format documents get turned into citations, even more specifically, how we populate the origin Solr field.

#5 - 2018-07-17 23:06 - Dave Vieglais

Metadata extraction and index population rules are documented at:

indexer-documentation.readthedocs.io

with the ISO TC211 rules under http://indexer-documentation.readthedocs.io/en/latest/generated/proc_isotc211Subprocessor.html

It looks like the solr index origin field is populated with:

```
//gmd:identificationInfo/gmd:MD_DataIdentification/  
gmd:citation/gmd:CI_Citation/  
gmd:citedResponsibleParty/gmd:CI_ResponsibleParty[  
gmd:role/gmd:CI_RoleCode/text() = "owner" or  
gmd:role/gmd:CI_RoleCode/text() = "originator" or  
gmd:role/gmd:CI_RoleCode/text() =  
"principalInvestigator" or gmd:role/gmd:CI_RoleCode/  
text() = "author"]/gmd:individualName/*/text()  
|  
//gmd:identificationInfo/gmd:MD_DataIdentification/  
gmd:citation/gmd:CI_Citation/  
gmd:citedResponsibleParty/gmd:CI_ResponsibleParty[  
(gmd:role/gmd:CI_RoleCode/text() = "owner" or  
gmd:role/gmd:CI_RoleCode/text() = "originator" or  
gmd:role/gmd:CI_RoleCode/text() =  
"principalInvestigator" or gmd:role/gmd:CI_RoleCode/  
text() = "author") and (not(gmd:individualName) or  
gmd:individualName[@gco:nilReason = "missing"])]/  
gmd:organisationName/*/text()
```