

Infrastructure - Decision #8189

Proposal to change the roles mapped to the origin Solr field for ISO docs

2017-10-02 18:04 - Bryce Mecum

| | | | |
|--|-------------|------------------------|-------------------|
| Status: | New | Start date: | 2017-10-02 |
| Priority: | Normal | Due date: | |
| Assignee: | Bryce Mecum | % Done: | 0% |
| Category: | | Estimated time: | 0.00 hour |
| Target version: | | | |
| Milestone: | None | | |
| Description | | | |
| <p>While discussing changing the behavior of the origin field in the ISO indexing component (https://redmine.dataone.org/issues/8165) to make it more selective about where in the document originators are pulled, Matt Jones (over email) suggested we revisit the set of roles as well. Let's do that in this Issue.</p> <p>The current set of roles mapped to the origin field are:</p> <ul style="list-style-type: none">• <i>originator</i>: party who created the resource• <i>author</i>: party who authored the resource• <i>owner</i>: party that owns the resource• <i>principalInvestigator</i>: key party responsible for gathering information and conducting research <p>This current set of roles may be surprising to some/many users so a possible outcome of this Issue is to greatly improve the content in our search index. This would have impacts on the CN and MNs running Metacat.</p> <p>Key points:</p> <ul style="list-style-type: none">• Matt's proposal is to exclude principalInvestigator from this list• The Research Workspace Member Node appears to be using the principalInvestigator role for one or more persons they want in their citation so if we follow Matt's proposal we may need to discuss this with them• I would lobby for only including originator and author but my reading of the definitions is a naïve one <p>I'd like us to have a discussion on this, make the relevant change to the codebase, and then bring the discussion back to the MN operators.</p> <p>Relevant links:</p> <ul style="list-style-type: none">• http://www.isotc211.org/2005/resources/Codelist/gmxCodeLists.xml (official? definitions for CI_RoleCode)• https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19115_and_19115-2_CodeList_Dictionaries (NOAA wiki entry for the role codes) | | | |
| Related issues: | | | |
| Related to Infrastructure - Task #8222: reindex all isotc211 content in produ... | | New | 2017-11-22 |
| Blocked by Infrastructure - Story #8504: Support creation of data citation re... | | New | 2018-03-19 |

History

#1 - 2017-11-03 01:04 - Bryce Mecum

I've done a lot of Googling and some emailing/chatting and found some useful references.

The overall question is: *"Who should show up in the data citation for an ISO 19115 record?"* Other considerations are:

- When people are authoring ISO metadata, are they thinking specifically about how their record will be rendered into a citation. If so, what's their interpretation?
- What will surprise users the least. We want to minimize surprise ("Why is this person listed in the citation?")
- We may need to balance *best practice* (whatever that is) with *community practice*

Re: Best practice

I essentially had no clue here prior to digging in. Data citations and citations in general are a total quagmire. I'm hearing and seeing essentially two things:

1. Most likely: Data citations should include the roles that logically map onto the idea of 'creator' and we need to decide on which roles we want to

include.

2. Less likely: Perhaps data citations should be constructed from all citedResponsibleParty listed in the ISO Citation, regardless of role

For (1), the only official-ish source I can find is on the NOAA wiki (

https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19115_and_19115-2_CodeList_Dictionaries#CI_RoleCode):

| ISO | DataCite |
|-----------------------|----------|
| author | creator |
| custodian | |
| distributor | |
| originator | creator |
| owner | |
| pointOfContact | |
| principalInvestigator | creator |
| processor | |
| publisher | |
| resourceProvider | creator |
| sponsor | |
| user | |

From that list, we see the cited roles should be author, originator, principalInvestigator, and resourceProvider.

For (2):

The ISO 19115 workbook, which is a human-readable manual for the standard, defines the element 'gmd:citation' as:

citation – Citation for the dataset

and its type, CI_Citation as:

CI_Citation – Bibliographic information to reference the resource.

My naïve reading of this is that all of the citedResponsibleParty contained inside the gmd:citation should (could?) be used in the citation for the data which would lead me use *all roles*.

The NOAA wiki for ISO_Citations (https://geo-ide.noaa.gov/wiki/index.php?title=ISO_Citations) says:

CI_Citation serves two purposes in the ISO 19115 Standard. First, it gives the information required to cite the data or the service (the resource) that is being described in the metadata. This CI_Citation can be part of the gmd:MD_DataIdentification or srv:ServiceIdentification objects.

The FORCE11 Data Citation Principles declaration (<https://www.force11.org/group/joint-declaration-data-citation-principles-final>) says:

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data[2].

This last point seems to making the case that either citations shouldn't just be the creators or that more people should be considered creators. I'm not sure which is the more sensible interpretation.

Re: Community practice

I'm not really familiar enough with the ISO metadata community to know where to even start trying to establish community practice. The only example of generating a single-line data citation from an ISO record comes out of . Help from the community would be greatly appreciated.

#2 - 2017-11-22 19:03 - Rob Nahf

- Related to Task #8222: reindex all isotc211 content in production to reflect final decisions from origin field mapping added

#3 - 2017-12-12 17:55 - Monica Ihli

- File ISO_19115 (copy)_page042.png added

- As shown in the attached diagram (From iso19115:2003 document) CI_Citation package describes entities CI_Citation and CI_ResponsibleParty.
- citedResponsibleParty contains zero or more instances of CI_ResponsibleParty.
- role value with CI_ResponsibleParty populated from an enumeration of values which may be found at: <http://www.isotc211.org/2005/resources/Codelist/gmxCodeLists.xml>
- /CI_Citation/citedResponsibleParty/CI_ResponsibleParty/role/CI_RoleCode

#4 - 2017-12-20 01:28 - Bryce Mecum

Had a great conversation yesterday with a NOAA researcher (Josh London, NMML) and was able to get their feedback on this topic. After sharing all the relevant details, he recommended a solution I hadn't even considered: Only map roles author and coAuthor into the origin field. This is probably quite different to what others want but I think he makes a great point. At the very least, it's what he finds least surprising.

Another important point he made is that, from a researcher perspective, it'd be nice to know two things: (1) what information you enter will go into the citation and (2) how I can control the order. These are separate concerns to this ticket but totally related and I at least wanted to write them down somewhere.

#5 - 2018-03-19 21:59 - Dave Viegla

- Blocked by Story #8504: Support creation of data citation record from solr record added

Files

| | | | |
|------------------------------|--------|------------|-------------|
| ISO_19115 (copy)_page042.png | 290 KB | 2017-12-12 | Monica Ihli |
|------------------------------|--------|------------|-------------|