

Infrastructure - Bug #8043

The origin field for EML documents isn't properly extracted when references are used

2017-03-10 21:36 - Bryce Mecum

| | | | |
|-------------------------|------------|------------------------|------------|
| Status: | New | Start date: | 2017-03-10 |
| Priority: | Normal | Due date: | |
| Assignee: | Jing Tao | % Done: | 0% |
| Category: | d1_indexer | Estimated time: | 0.00 hour |
| Target version: | CCI-2.4.0 | Story Points: | |
| Milestone: | None | | |
| Product Version: | | | |

Description

We just ran into this with the following EML record: <https://knb.ecoinformatics.org/#view/doi:10.5063/F15B00CC>

The EML has six creators (Kiesecker, Fargione, Baruch-Mordo, Trainor, Ryan, Patterson) but the origin field in the Solr index has two (Ryan, Patterson). After some digging, we realized this was likely because the indexing component responsible for EML doesn't respect EML references. The XML for the relevant section is:

```
<creator scope="document">
<references>1484778487589</references>
</creator>
<creator scope="document">
<references>1484778426939</references>
</creator>
<creator scope="document">
<references>1484778028081</references>
</creator>
<creator scope="document">
<references>1484778171131</references>
</creator>
<creator id="1485385283277" scope="document">
<individualName>
<salutation>Dr.</salutation>
<givenName>Joe</givenName>
<surName>Ryan</surName>
</individualName>
<organizationName>University of Colorado Boulder</organizationName>
<positionName>Professor</positionName>
<electronicMailAddress>joseph.ryan@colorado.edu</electronicMailAddress>
</creator>
<creator id="148477776976" scope="document">
<individualName>
<salutation>Dr.</salutation>
<givenName>Lauren</givenName>
<surName>Patterson</surName>
</individualName>
<organizationName>Duke University</organizationName>
<positionName>Water Policy Associate</positionName>
<address scope="document">
<deliveryPoint>Nicholas Institute for Environmental Policy Solutions, Duke University</deliveryPoint>
<city>Durham</city>
<administrativeArea>NC</administrativeArea>
<postalCode>27708</postalCode>
<country>USA</country>
</address>
<electronicMailAddress>lauren.patterson@duke.edu</electronicMailAddress>
</creator>
```

It would be really nice if the origin field got populated with all those referenced creators.

History

#1 - 2017-03-10 21:39 - Bryce Mecum

Also, and very importantly, chris guesses that the indexing component isn't doing anything at all with references so other fields should be affected here too.

#2 - 2017-03-10 21:56 - Jing Tao

- Assignee set to *Jing Tao*

- Target version set to *CCI-2.4.0*

#3 - 2017-03-28 16:01 - Dave Vieglais

- Project changed from *CN Index* to *Infrastructure*

- Category set to *d1_indexer*

- Milestone set to *None*

#4 - 2019-02-12 18:45 - Matthew Jones

- Description updated