

Member Nodes - Story #8225

MNDeployment # 7895 (Operational): Pangaea

Customize Indexing & View for gmd-pangaea

2017-12-06 19:41 - Monica Ihli

Status:	In Progress	Start date:	2017-12-06
Priority:	Normal	Due date:	
Assignee:	Rob Nahf	% Done:	30%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			
Description			
An example metadata record: http://cn-sandbox.test.dataone.org/cn/v2/object/doi:10.1594/PANGAEA.877809_.201711172109			
This record in the search interface on sandbox: https://search-sandbox.test.dataone.org/#view/doi:10.1594/PANGAEA.877809_.201711172109			
Currently, alternate access point is pulling the link from: /ns0:MD_Metadata/ns0:distributionInfo[1]/ns0:MD_Distribution[1]/ns0:transferOptions[1]/ns0:MD_DigitalTransferOptions[1]/ns0:onLine[1]/ns0:CI_OnlineResource[1]/ns0:linkage[1]/ns0:URL[1]			
However, Pangaea wishes users to be directed towards a landing page where they are able to obtain METADATA in multiple formats, found in: /ns0:MD_Metadata/ns0:dataSetURI[1]/ns2:CharacterString[1]			
The landing page for this example: https://doi.pangaea.de/10.1594/PANGAEA.877809			
Related issues:			
Related to Infrastructure - Task #8499: Improve rendering of http://www.isotc...		New	2018-03-14
Related to Infrastructure - Task #8219: verify proper rendering of http://www...		Closed	2017-11-21

History

#1 - 2017-12-06 21:00 - Rob Nahf

- Tracker changed from Task to Story

The dataSetURI field is currently unparsed and could either be added to or replace the existing serviceEndpoint, or be parsed into another field in the index.

The usability issue here is whether or not automated workflows can get past the html and navigate to the dataset. So, replacing the value in the index make break some things for the user.

Also, probably not related: the landing page doesn't provide the data in multiple formats, but offers different formats for data citation (RIS, BibTeX, plain text). These formats don't seem to be encoded in the metadata itself, but the landing page links point to other Pangaea service endpoints (by adding "?format_citation=" to the url). The link to the data is still only to one file (the .xlsx)

#2 - 2018-01-04 20:45 - Monica Ihli

12/11 MNW call - delay indexing changes because the data can always be reindexed later.
12/19 Maint call - pushed back possible Pangaea index parsing tasks into 2.3.8.

So this ticket is on hold until later.

#3 - 2018-03-20 17:18 - Amy Forrester

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

#4 - 2018-03-26 14:12 - Amy Forrester

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

#5 - 2018-03-26 14:31 - Monica Ihli

- Description updated

#6 - 2018-03-26 19:03 - Amy Forrester

3/26/18: Dave to test allowing multiple access points to be listed in record

#7 - 2018-03-27 17:01 - Amy Forrester

- File *Pangaea issues.pdf* added

3/27/18: Issues from Michael Diepenbroek after *scanning through the harvested records*.

- PANGAEA – always capital letters, in : „Pangaea Data Publisher ...“ Correct is: „PANGAEA – Data Publisher ...“
- The DOI as part of the citation should be a hot link !! This is the fastest way to access the data. Also, why do you change the official DOI representation using <https://doi.org> ... by the not actionable prefix „doi: ..“
- „Version“ with DataONE „Id“ should not part of the citation (not correct!) – PANGAEA allows versioning with each version having a separate DOI. Old versions are interlinked with new version – visible in the metadata as reference.
- We are missing citation details. Some data sets have references included in the citation (supplementary articles) listed in „otherCitationDetails“.
- Citation on your site should be identical to what is listed on the splash page of PANGAEA.
- „Alternate data access“ is redundant with „Service Endpoint“ (for normal users hard to understand). Link is not clickable.
- „Id“ as part of metadata is confusing. At least it should be made clear that this is an internal DataOne ID (never to be used in citations!)
- „Origin“ and „Investigator“ seem redundant. In PANGAEA we differentiate between the authorship (in PANGAEA „cited responsible party“, role code „author“ not „origin“) of the data and the principal investigators (PI) (role code „principalInvestigator“) which are listed in the identification info. Authorship and PIs may not be identical. Where did you get the „Investigators“ from?
- „Contact Organisation“ could be a hot link for convenience.
- Listing the DOI of the data set as „Series ID“ below „Other“ is odd. A resolvable PID should replace the DataONE „Id“.

See Monica summary: <https://hpad.dataone.org/201803-MaintenanceStandup>

{Dave} All reasonable comments, and several are generally applicable to the search UI across other content besides PANGAEA. Addressing most of these issues will require UI changes, so will require some help from Chris and Lauren.

**discuss on Thursday (3/29) meeting with UI team

#8 - 2018-03-27 17:04 - Amy Forrester

- File deleted (Pangaea issues.pdf)

#9 - 2018-04-02 18:40 - Monica Ihli

Follow-Up Based on Meeting with Search UI Team on 3/29

- We can add hyperlinks to DOIs to the <https://doi.org/...> URL
- TODO: Add Pangea ISO formatID to the list of ISO formatIDs in view service
- We can probably add some additional indexing rule for the Pangea formatID that identifies the landing-page access point.
- We will try to explain to Pangea our reasonings for including the PID in the citation. (based on our handling of immutable content and for versioning).
- We can handle linking the DOI on the search result page. If it looks like a DOI, it will be formatted out to the DOI resolver. Lauren will open tickets.
- "Alternate Data Access" and "Service Endpoint" - explain our differentiation between human and machine readable access points.
- Regarding View: What they are seeing right now is fallback behavior for view service. Bryce will let me know when that is updated. If it is still an issue, we can index so that author role is used instead of origin.
- Possibly put together a conference call between major players in the community as to what constitutes an appropriate ISO mapped citation. (Chris & Monica). Get some examples used by community.

<https://hpad.dataone.org/MwNgJghgnKCMC0AGAZgVkfALBCZ4QGnMFkAjAg1CqRTSIA==>

#10 - 2018-04-02 19:26 - Amy Forrester

- Related to Task #8499: Improve rendering of <http://www.isotc211.org/2005/gmd-pangaea> in search UI added

#11 - 2018-04-02 19:27 - Amy Forrester

- Related to Task #8219: verify proper rendering of <http://www.isotc211.org/2005/gmd-pangaea> format metadata in search UI added

#12 - 2018-04-05 15:32 - Monica Ihli

DOI linking ticket created in GitHub for Metacat: <https://github.com/NCEAS/metacatui/issues/544>

View changes documented in ticket: <https://github.com/NCEAS/metacat/issues/1232>

#13 - 2018-04-30 19:03 - Monica Ihli

• PANGAEA – always capital letters, in : „Pangaea Data Publisher ...“ Correct is: „PANGAEA – Data Publisher ...“
(FIXED)

• The DOI as part of the citation should be a hot link !! This is the fastest way to access the data. Also, why do you change the official DOI representation using <https://doi.org> ... by the not actionable prefix „doi: ..“

(Hi @lauren, we are reviewing the current status of pangaea and I believe that you were working on the DOI being hyperlinked. I see that this is the case in stage as of now (https://search-stage.test.dataone.org/#view/doi:10.1594/PANGAEA.802166_201711191956) in the record details. Is there any plan to incorporate that into how a doi is displayed in the data citation, or is that outside the scope of the work you were going to perform?)

• „Version“ with DataONE „Id“ should not part of the citation (not correct!) – PANGAEA allows versioning with each version having a separate DOI. Old versions are interlinked with new version – visible in the metadata as reference.

(We are in process of displaying SID only for pangaea formatid which will result in the seriesID (DOI) being displayed only as part of the citation.

However, at this time, it would have to be a future feature addition to display where multiple versions of the same object can be found. This has been under discussion but there are no immediate plans to incorporate it into current development at this time)

- We are missing citation details. Some data sets have references included in the citation (supplementary articles) listed in „otherCitationDetails“. Citation on your site should be identical to what is listed on the splash page of PANGAEA.

(We are currently doing a complete review of how data citation is constructed in the DataONE system. If we are able to accommodate that change to the data citation specifically, we will let you know. In the meantime, this information is still available in the complete record view.)

- „Alternate data access“ is redundant with „Service Endpoint“ (for normal users hard to understand). Link is not clickable.

(The Service Endpoints is for designated machine readable consumption of what's on the other end. It's not actually redundant because the access point is meant for humans. However, we are relabeling the alternate data access as "Distribution" and making it clickable instead of a copy button.)

- „Id“ as part of metadata is confusing. At least it should be made clear that this is an internal DataOne ID (never to be used in citations!)

(Addressed already by not including dataone PID in the citation)

- „Origin“ and „Investigator“ seem redundant. In PANGAEA we differentiate between the authorship (in PANGAEA „cited responsible party“, role code „author“ not „origin“) of the data and the principal investigators (PI) (role code „principalInvestigator“) which are listed in the identification info. Authorship and PIs may not be identical. Where did you get the „Investigators“ from?

(Please see revised view of full record, which does show redundant information in the same way)

- „Contact Organisation“ could be a hot link for convenience.

(Contact information and organizational information hyperlinked where available)

- Listing the DOI of the data set as „Series ID“ below „Other“ is odd. A resolvable PID should replace the DataONE „Id“.

(Already addressed)

#14 - 2018-05-02 17:49 - Monica Ihli

- *Subject changed from Customize Indexing for gmd-pangaea to Customize Indexing & View for gmd-pangaea*

#15 - 2018-05-02 17:49 - Monica Ihli

We have a set of changes and need to manage them:

1. Making links in the citation clickable. Lauren did this work [codebase: MetacatUI 1.14.15]
2. Making Pangaea's citations not show their PID. I did this work, can be hot patched ASAP at the *same time* as (1) [codebase: MetacatUI 1.14.15]
3. Removed Alternate Data Access Table from MetacatUI [codebase: MetacatUI 1.14.15]
4. Improvements to Pangaea's (and all ISO producer's) metadata. *Depends on 3* [codebase: Metacat, can be hot-patched or we can do a patch release, depending on @jing's thoughts]

(3) and (4) need to happen in tandem

#16 - 2018-05-11 00:03 - Amy Forrester

note from Monica (5/10/18): some of the DOIs hyperlinked will not resolve, but that seem to be because they have not yet completed activation of the DOI. Example: <https://search.dataone.org/index.html#view/c7ded2e1a46d85d7adb682d0caa2f074> on our site is <https://doi.pangaea.de/10.1594/PANGAEA.889138> on theirs, where they indicate in their citation that DOI registration is in progress. There's not much we can do about that.

#17 - 2018-05-17 14:30 - Monica Ihli

(1)

PANGAEA - always capital letters, in : „Pangaea Data Publisher ...“ Correct is: „PANGAEA - Data Publisher ...“

- Fixed - Displays as PANGAEA now in search UI of production.

(2)

The DOI as part of the citation should be a hot link !! This is the fastest way to access the data. Also, why do you change the official DOI representation using <https://doi.org> ... by the not actionable prefix doi

- The DOI is now hyperlinked in the citation.

(3)

Version“ with DataONE „Id“ should not part of the citation (not correct!) - PANGAEA allows versioning with each version having a separate DOI. Old versions are interlinked with new version - visible in the metadata as reference.

- The SID (DOI) only is now displayed for Pangaea. This is specific to pangaea formatId only.
- See issue <https://redmine.dataone.org/issues/8574>.
- Note also the related sub-task [#8576](#).

(4)

We are missing citation details. Some data sets have references included in the citation (supplementary articles) listed in „otherCitationDetails“. Citation on your site should be identical to what is listed on the splash page of PANGAEA.

- We are currently doing a complete review of how data citation is constructed in the DataONE system. If we are able to accommodate that change to the data citation specifically, we will let you know. In the meantime, this information is still available in the complete record view.

(5)

Alternate data access“ is redundant with „Service Endpoint“ (for normal users hard to understand). Link is not clickable.

- The Service Endpoints is for designated machine readable consumption of what's on the other end. It's not actually redundant because the access point is meant for humans.
- However, we have relabeled the "alternate data access" section as "Distribution" and making it clickable instead of a copy button.

(6)

Id“ as part of metadata is confusing. At least it should be made clear that this is an internal DataOne ID (never to be used in citations!)

- This concern is addressed in that we no longer include the DataONE PID in the citation

(7)

„Origin“ and „Investigator“ seem redundant. In PANGAEA we differentiate between the authorship (in PANGAEA „cited responsible party“, role code „author“ not „origin“) of the data and the principal investigators (PI) (role code „principalInvestigator“) which are listed in the identification info. Authorship and PIs may not be identical. Where did you get the „Investigators“ from?

- Please see revised view of full record, which provides a slimmer, less redundant view of the information.

(8)

„Contact Organisation“ could be a hot link for convenience.

- Contact information and organizational information are hyperlinked where available.

(9)

Listing the DOI of the data set as „Series ID“ below „Other“ is odd. A resolvable PID should replace the DataONE „Id“

- Already addressed.

Additional:

An additional improvement is that parameters are now parsed from the iso metadata and displayed in a table of attributes similar to as displayed on pangea's website. Example:

<https://search.dataone.org/index.html#view/b55c4390a74294db8306ae8ce5bfd653>