

EDIT - feature request #10271

Implement CoL-DP export

03/13/2023 05:27 PM - Andreas Müller

Status:	In Progress	Start date:	
Priority:	Priority14	Due date:	
Assignee:	Andreas Müller	% Done:	30%
Category:	cdmadapter	Estimated time:	0:00 hour
Target version:	Release 5.42		
Severity:	normal		
Description			
<p>Important for exports to Col and GCB and maybe also WFO. Alternatively implement it as DwC-A backbone export (which is currently the only format supported by WFO, but Roger H. mentioned that and import will be implemented once data in this format exists).</p> <p>https://github.com/CatalogueOfLife/coldp</p> <p>Short titles for nom. refs. will be added to COL-DP: https://github.com/CatalogueOfLife/coldp/issues/74</p>			
Related issues:			
Related to EDIT - feature request #10273: WFO content export (DwC-A)		In Progress	
Related to EDIT - feature request #10446: WFO Backbone export		Resolved	
Related to EDIT - feature request #10449: Allow filtering out synonyms in lis...		Closed	
Related to EDIT - feature request #10451: Allow adding fullName (name+author)...		Resolved	
Precedes EDIT - feature request #10384: Add col-dp export to exports in taxed...		Closed	

Associated revisions

Revision 1cb74dfd - 07/19/2023 11:07 PM - Andreas Müller

ref #10271 first incomplete version of COL-DP export

Revision 82983a46 - 07/20/2023 09:21 AM - Andreas Müller

ref #10271 fix ClearDB datasets

Revision a2428d57 - 07/20/2023 09:41 AM - Andreas Müller

ref #10271 fix ClearDB datasets (cont.)

Revision dce70ce8 - 07/20/2023 10:30 AM - Andreas Müller

ref #10271 fix ClearDB datasets (cont. II)

Revision 25877fda - 07/20/2023 02:28 PM - Andreas Müller

ref #10271 fix ClearDB datasets (cont. III)

Revision 9f3695e5 - 07/20/2023 04:04 PM - Andreas Müller

ref #10271 add common base classes for CDM light and COL-DP export

Revision df3e9a0e - 07/21/2023 02:35 PM - Andreas Müller

ref #10271 further data types for COL-DP and clean up COL-DB and CDM light

Revision 4f9e9b3f - 09/04/2023 10:41 AM - Katja Luther

ref #10271: Transformer needs to be serializable

Revision a11f9efc - 09/05/2023 10:57 AM - Andreas Müller

ref #10271 remove unnecessary parameters from COL-DP export configurator

Revision c7872fea - 09/05/2023 11:16 AM - Andreas Müller

ref #10271 remove unnecessary parameters from COL-DP export configurator (cont.) and set createZipFile to true

Revision e80c6b90 - 10/26/2023 09:30 PM - Andreas Müller

ref #10271 some improvements to COL-DP Export

Revision c5ec2ffc - 10/26/2023 09:51 PM - Andreas Müller

ref #10271 some improvements to COL-DP Export (fix export type test in CDM light)

Revision a9079a6e - 01/09/2024 10:47 PM - Andreas Müller

ref #10271 add error handling and avoid NPE for unhandled name relationship types

Revision 81cfaeaa - 01/09/2024 10:50 PM - Andreas Müller

ref #10271 avoid NPE for media without mediaCreated

Revision 82a98476 - 02/01/2024 11:11 PM - Andreas Müller

ref #10271 fix protectd titleCache handling in ColDP Export

Revision 89a9b872 - 02/01/2024 11:12 PM - Andreas Müller

ref #10271 add IPNI normalizeAuthor to ColDP export

Revision 404beffb - 02/12/2024 11:38 PM - Andreas Müller

ref #10271 handle ipni, tropicos and index fungorum identifiers

Revision ce907050 - 02/13/2024 01:50 PM - Andreas Müller

ref #10271 handle CURTI identifiers separately

Revision 262ee0f0 - 02/15/2024 12:00 PM - Andreas Müller

ref #10271 remove replacing of "wfo-" in WFO Identifiers (according to documentation change)

History

#1 - 03/15/2023 12:07 PM - Andreas Müller

- Related to feature request #10273: WFO content export (DwC-A) added

#2 - 03/15/2023 12:11 PM - Andreas Müller

- Description updated

#3 - 03/15/2023 03:17 PM - Andreas Müller

- Description updated

#4 - 07/19/2023 07:15 PM - Andreas Müller

Questions:

- how to map reference type article? "article" or "article journal"?
- how to handle nomencl. refernces in the reference table as they usually do not have full titles and authorship differs => there are 3 new fields now (see above).
- how to handle details/microReferences, e.g. for nom. refs.

- ref.volume, .issue and .edition that are not numbers (and .issue in general as we don't split this)
- is distribution status "domesticated" relevant for plants? (<http://api.checklistbank.org/vocab/distributionstatus>)
- should doubtfully native be included in "native" and should "unknown endemism" be include in native?
- what to use for the nomenclatural code? The name or the acronym? (<http://api.checklistbank.org/vocab/nomCode>)

#5 - 07/20/2023 08:20 PM - Andreas Müller

- Status changed from New to In Progress

- Priority changed from New to Highest
- Target version changed from Release 5.45 to Release 5.40
- % Done changed from 0 to 30

#6 - 07/27/2023 01:07 PM - Andreas Müller

- Description updated

#7 - 07/27/2023 01:10 PM - Andreas Müller

WGB:

I looked at the CoLDP documentation and I cc to Markus (who is on vacation I believe).

Reference.type: I think our articles are always journal articles and should be mapped to this term (in Zotero there is no generic "article"). If there would be a lone newspaper or magazine article somewhere in our data, it wouldn't matter, because the format is practically the same.

nomenclatural references:

I don't think that differing authorship is a problem, because the nomenclatural author is handled by the Name table: Name.authorship. Here the standard abbreviated author should be entered for botanical names.

The reference detail is also handled there: Name.publishedInPage.

Having the Name.publishedInYear here as well refers more to zoological names, but should be filled for botany as well, because it can be useful to distinguish isonyms etc.

Indeed, the rest of the citation has to be handled in the Reference table (this is also true for the simplified NameUsage table).

I see no way here to have bibliographical and nomenclatural references for botany represented in a single record – the standard abbreviations cannot be entered in parallel to the full title. So separate records have to be created. (No title abbreviation in BibTex neither, I believe).

The relationship between nomenclatural record and bibliographic record that we support in the CDM is thus lost.

I see no tricks to get around that - since there is no recursive relationship for the Reference table ("containerID"), this cannot be solved in the current CoLDP model. However, avoiding recursive relationships in an exchange model definitely helps.

@Markus: This could be overcome by a TEXT field in Name for nomenclaturalReference or in Reference for nomenclaturalTitleAbbreviation. The latter would be more normalised, the former would have the advantage to have all name data in the Name table.

MD:

just quick from Cologne: CoLDP offers 3 ways to share structured references, CSV, BibTex or CSL-JSON.

CSL-JSON is extensible and has many fields that can be used, including "short" titles like title-short to hold an abbreviated version:

<https://github.com/citation-style-language/schema/blob/master/schemas/input/csl-data.json>

It is supported by most of the large reference information systems like CrossRef, DataCite, etc and you can resolve any DOI to it easily.

Providing JSON is rather difficult for most people though, so I have created an issue to extend the CSV format to also be able to do the same, also for the author and journal: <https://github.com/CatalogueOfLife/coldp/issues/74>

If you want use that already. It will take some time to be properly supported, but should happen some time this year for sure. It would be a shame to have 2 references for the same thing.

As Walter said the Name.publishedInYear is useful for botanical and zoological names. Especially for botany it allows to compare and sort names on their year of publication, which is otherwise not existing in the authorship. And not in all cases a reference record exists.

#8 - 08/29/2023 10:18 AM - Andreas Müller

- Precedes feature request #10384: Add col-dp export to exports in taxeditor added

#9 - 09/07/2023 06:54 PM - Andreas Müller

- Target version changed from Release 5.40 to Release 5.41

#10 - 10/27/2023 11:21 PM - Andreas Müller

- Target version changed from Release 5.41 to Release 5.42

#11 - 01/01/2024 03:38 PM - Andreas Müller

- Related to feature request #10446: WFO Backbone export added

#12 - 01/12/2024 05:14 PM - Andreas Müller

- Related to feature request #10449: Allow filtering out synonyms in list exports added

#13 - 01/12/2024 05:45 PM - Andreas Müller

- Related to feature request #10451: Allow adding fullName (name+author) to coldp name table added

#14 - 02/01/2024 10:44 AM - Andreas Müller

- Priority changed from Highest to Priority14