

EDIT - task #10324

Export Caucasus data

05/23/2023 01:26 PM - Andreas Müller

Status: Closed	Start date:
Priority: Priority13	Due date:
Assignee: Andreas Müller	% Done: 100%
Category: cdmadapter	Estimated time: 0:00 hour
Target version:	
Severity: normal	

Description

... to Georgia, Armenia and Azerbaijan databases.

Filters:

- Taxonomic: ~~all taxa, published and unpublished, but no Kew/ILDIS data~~
- Geographic: ~~all taxa with distribution in the respective countries~~
 - ~~... and distributions only from the respective countries~~
 - ~~include parent taxa (without distribution) and absent status~~
 - ~~status undefined (currently imported, but not shown)~~
- Content: ~~distribution and common names~~, but common name filter still needs to be discussed (which languages and which areas - see also [#10324#note-28](#)), exclude other features(?)
- Feature tree
- Status: ~~do not import aggregated descriptions and empty (after filter) descriptions~~
- Endemism
 - fine tuning
- Area trees (including Caucasus local areas), excluding common name areas not in use
- Name duplicates (synonymy <-> name used in source): sent table to ERS (2023-07-04)
- ~~orphaned taxa introduced due to name reusage, e.g. -~~
https://portal.cybertaxonomy.org/georgia/cdm_dataportal/taxon/6ee1e37a-0ec7-4930-bbc2-0d3e7cc80eb8
- Decision on Tcs/Cc Taxa (see TCS_CS_Taxa.xlsx)

Before final import:

- check status of Kew taxa with ERS

After import:

- rename Common Name -> Common Names
- Update taxon treeindex

Related issues:

Related to EDIT - task #10239: Caucasus databases and portals available	Closed	
Related to EDIT - feature request #9197: Handle caucasus conspectus data in E...	Rejected	
Related to EDIT Platform Etablierung - task #8689: Export Flora of Greece Bup...	Closed	11/18/2019
Related to EDIT - task #10183: Import Buxales via Cdm2Cdm	Closed	
Related to EDIT - feature request #9771: Implement Cdm2Cdm for vocabularies	Closed	

Associated revisions

Revision 9c23eca2 - 06/14/2023 10:11 PM - Andreas Müller

ref #10324 add destinations and activator for caucasus exports

Revision 12a9060e - 07/02/2023 12:29 PM - Andreas Müller

ref #10324 add latest changes for Caucasus export

Revision bcb3b9a1 - 07/02/2023 12:45 PM - Andreas Müller

ref #10324 adaptations to Cdm2CdmImport for Caucasus exports

Revision 1f8b88d2 - 07/08/2023 12:33 PM - Andreas Müller

ref #10324 fix namedArea fetching in import base

Revision d2d3d57e - 07/13/2023 10:44 AM - Andreas Müller

ref #10324 add endemism handling for Cdm2CdmImport

Revision aafb3ba2 - 07/13/2023 10:44 AM - Andreas Müller

ref #10324 fix taxonName->taxonBase propagation in Cdm2CdmImport

Revision 864a136d - 07/13/2023 10:44 AM - Andreas Müller

ref #10324 allow absent terms for geographic TaxonNodeFilter

Revision 498e4267 - 07/26/2023 06:48 PM - Andreas Müller

ref #10324 some changes on EuroMedCaucasusChecklistExportActivator and add euromed-caucasus as source

Revision 72c12ce0 - 09/21/2023 10:35 AM - Andreas Müller

ref #10324 add source to endemics

Revision e327e364 - 09/21/2023 10:35 AM - Andreas Müller

ref #10324 fix mapping for nomenclatural status

Revision b8cf681f - 09/21/2023 10:36 AM - Andreas Müller

ref #10324 add filter for heterotypic synonyms

Revision 7731f1ed - 09/28/2023 09:06 PM - Andreas Müller

ref #10324 add synonym filter and other changes for final Georgia import

Revision d24ce5cf - 01/01/2024 03:45 PM - Andreas Müller

ref #10324 Adapt CaucasusExportActivator to latest changes

Revision 2adc43b9 - 01/28/2024 10:13 AM - Andreas Müller

ref #10324 update childrenCount correctly in Cdm2CdmImportBase

History

#1 - 05/23/2023 01:26 PM - Andreas Müller

- Related to task #10239: Caucasus databases and portals available added

#2 - 05/23/2023 01:28 PM - Andreas Müller

- Related to feature request #9197: Handle caucasus conspectus data in E+M and caucasus portal added

#3 - 05/23/2023 01:50 PM - Andreas Müller

- Related to task #8689: Export Flora of Greece Bupleurum using Cdm2Cdm export added

#4 - 05/23/2023 01:51 PM - Andreas Müller

- Related to task #10183: Import Buxales via Cdm2Cdm added

#5 - 05/23/2023 01:51 PM - Andreas Müller

- Related to feature request #9771: Implement Cdm2Cdm for vocabularies added

#6 - 05/23/2023 08:45 PM - Andreas Müller

What about endemism data? See also [#10239#note-2](#)

#7 - 05/23/2023 08:48 PM - Andreas Müller

Also discuss WFO-IDs for names

NaK:

die WFO-IDs sollten in die Kaukasus-Instanzen auf jeden Fall rein.

#8 - 06/14/2023 11:13 PM - Andreas Müller

- Status changed from New to In Progress

- % Done changed from 0 to 10

#9 - 06/16/2023 09:36 AM - Andreas Müller

ERS:

we did not discuss endemism yesterday. Data on endemism will be important for each checklist.

Which data on endemism do we currently have in Euro+Med Plantbase?

Endemism in Euro+Med Plantbase is not an attribute for a country or other subarea, but for the top area Euro+Med. If we have for a given taxon the area Euro+Med and the indication endemic, and distribution records for only one country, then this taxon is endemic to that country.

I would suggest, but this needs to be discussed with the partners, that two (or even three) categories of endemics will be useful for each country:

- (local endemics (at subarea level, e.g. Nakhchivan or Abkhazia))
- country endemics (Ab, Ar or Gg)
- Caucasus endemics (occurring in more than one country, but only in the Caucasus, i.e. only Ab, Ar, Gg or Rf(CS)=North Caucasus).

So we need to decide how to handle endemism in the three new instances, and which endemism data should be included in the export from E+M.

#10 - 06/16/2023 09:51 AM - Andreas Müller

MM:

Georgia is reported among distribution for *Abies nordmanniana* (Steven) Spach without reference. The references are added in TaxEditor but it seems they are not published in Euro+Med.

I am reporting the cases that I think may be they need to be edited in Euro+Med. Please let me know if these are normal and do not need to be reported.

ERS:

thank you for observing this. Both *Cystopteris fragilis* and *Abies nordmanniana* are taxa with subspecies, and you find the typical subspecies in Georgia. The reference is only given for the subspecies, not for the species. The distribution for the species is calculated from the subspecies, and because it is calculated, the reference is not given directly (but we have the possibility to add it if we want). So you do not need to report further cases, this is a known fact to us that data here are not always complete.

Maybe this is also something we need to think about when creating the export files for the three Caucasus instances?

#11 - 06/27/2023 05:23 PM - Andreas Müller

Available languages for common names:

```
SELECT deb.language_id, lang.titleCache, lang.uuid, COUNT(*) AS n
FROM DescriptionElementBase deb INNER JOIN DefinedTermBase lang ON lang.id = deb.language_id
GROUP BY deb.language_id
ORDER BY lang.titleCache
```

#12 - 06/27/2023 05:29 PM - Andreas Müller

- Description updated

#13 - 06/27/2023 06:06 PM - Andreas Müller

- Description updated

#14 - 06/28/2023 12:00 AM - Andreas Müller

Remove import sources does not work yet

#15 - 06/28/2023 06:55 AM - Andreas Müller

Excluded nodes:

```
SELECT tn.id, tn.UUID, treeIndex, n.nameCache
FROM TaxonNode tn INNER JOIN TaxonBase tb ON tb.id = tn.taxon_id INNER JOIN TaxonName n ON n.id = tb.name_id
WHERE (1=0) OR
tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#50717#51601#40724#' -- (Amaryllidaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52420#48555#40499#' -- (Araliaceae)
// OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#50717#51601#40629#' -- (Asparagaceae) -- mostly done
so I removed it from negative filter
```

OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52436#51035#44434#' -- (Betulaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52436#52441#42193#' -- (Euphorbiaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52436#50871#44688#' -- (Fabaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52436#51035#39628#' -- (Fagaceae)

OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#50717#51601#40557#' -- (Iridaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#50717#51116#49935#' -- (Juncaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#52420#49316#39608#' -- (Lamiaceae)
OR tn.treeIndex LIKE '#t10#10#56284#54730#51914#54954#50717#51601#40684#' -- (Orchidaceae)
ORDER BY tn.treeIndex

#16 - 06/28/2023 05:48 PM - Andreas Müller

- Description updated

#17 - 06/28/2023 05:48 PM - Andreas Müller

- Description updated

#18 - 06/28/2023 10:26 PM - Andreas Müller

- Description updated

- % Done changed from 10 to 30

#19 - 07/01/2023 11:27 AM - Andreas Müller

- Description updated

#20 - 07/01/2023 11:31 AM - Andreas Müller

- Description updated

#21 - 07/01/2023 11:33 AM - Andreas Müller

- Description updated

#22 - 07/02/2023 12:05 PM - Andreas Müller

- % Done changed from 30 to 50

#23 - 07/02/2023 12:05 PM - Andreas Müller

- Description updated

#24 - 07/02/2023 12:06 PM - Andreas Müller

- Description updated

#25 - 07/04/2023 03:48 PM - Andreas Müller

- Description updated

#26 - 07/04/2023 10:40 PM - Andreas Müller

- Description updated

#27 - 07/04/2023 10:45 PM - Andreas Müller

- Description updated

#28 - 07/04/2023 10:58 PM - Andreas Müller

Common names. ERS:

Also, wenn man individuell filtern kann, sollte es eher so aussehen:

Georgia: - Georgian (Georgia)
- Russian (Georgia)
- Russian (Russia)

Armenia: - Armenian (Armenia)
- Russian (Armenia)
- Russian (Russia)

Azerbaijan: - Azerbaijani (Azerbaijan)

- Russian (Azerbaijan)
- Russian (Russia)

#29 - 07/04/2023 10:59 PM - Andreas Müller

- Description updated

#30 - 07/04/2023 11:18 PM - Andreas Müller

- Description updated

#31 - 07/07/2023 01:06 PM - Andreas Müller

- Target version changed from Release 5.44 to Release 5.40

#32 - 07/08/2023 09:35 PM - Andreas Müller

- Description updated

#33 - 07/10/2023 09:10 AM - Andreas Müller

- Description updated

#34 - 07/10/2023 01:10 PM - Andreas Müller

- Priority changed from New to Highest

- Target version changed from Release 5.40 to Release 5.44

#35 - 07/12/2023 05:04 PM - Andreas Müller

Check how many taxa are already imported:

```
SELECT tn.id, tn.created, tn.countChildren, tn.treeIndex, tn.parent_id, tn.sortIndex, acc.id tId, acc.titleCache
FROM TaxonNode tn INNER JOIN TaxonBase acc ON acc.id = tn.taxon_id
ORDER BY tn.treeIndex, acc.titleCache;
```

#36 - 07/12/2023 06:19 PM - Andreas Müller

- Description updated

#37 - 07/31/2023 05:40 PM - Andreas Müller

ERS:

Specifications, as discussed during the workshop in Georgia:

- No common names in Russian nor in English. Georgian common names will be provided by our counterparts for a data import.
- No heterotypic synonyms, because deleting superfluous synonyms will be much more cumbersome than adding only the needed ones
- Only taxa with distribution in Gg, Gg(A), Gg(D), Gg(G)
- Only Euro+Med Taxa (no ILDIS or Kew Taxa)

A table of administrative subdivisions of Georgia, as well as shape files, will be provided by our counterparts in Georgia. I believe this can also be added after the instance split.

At present, additional data (mainly names and distributions for grasses) are input in the E+M main instance, because they are needed for all three countries. This is planned until end of August.

Georgians are not happy with the present E+M subdivision of Georgia in Abkhazia (Gg(A)), Adjara (Gg(D) and the rest of Georgia (Gg(G)), because this somehow implies that those areas are on the same administrative level, which is not the case. For them, Georgia is Georgia and should not be divided in E+M, but only in the Georgian instance, where there will be ca. 9-10 subdivisions, Abkhazia and Adjara are just two of them and extant data for them will be kept as such only in the Georgian slice.

However, after the definite export is done, the Georgian subdivisions in Euro+Med Plantbase should be merged and not used nor shown anymore.

I believe that the Georgian colleagues have a point here, because subdivisions are used somewhat inconsistent in E+M: For instance, Hs(A) is not a part of Spain, but independent Andorra, Si(M) is not part of Sicily, but independent Malta, Au(L) is not part of Austria, but independent Liechtenstein and so on – so one might believe that also Gg(A) and Gg(D) are not part of Georgia, but independent countries. On the other hand, Rf(C), Rf(E) and other subdivisions of Russia are certainly part of Russia, but those subdivisions do not correspond to administrative units. So at present, Gg(A) and Gg(D) are the only administrative subdivisions which are not independent countries. So we have good reason not to use them in future and merge the data for them into Georgia as a whole.

Please come back to me with any questions regarding the export for Georgia.

#38 - 09/07/2023 06:58 PM - Andreas Müller

- Target version changed from Release 5.44 to Release 5.41

#40 - 10/27/2023 11:21 PM - Andreas Müller

- Target version changed from Release 5.41 to Release 5.42

#41 - 01/10/2024 11:06 AM - Andreas Müller

- Status changed from In Progress to Resolved

are there open issues?

#42 - 01/11/2024 05:08 PM - Andreas Müller

- Description updated

#43 - 01/17/2024 08:44 PM - Andreas Müller

- Priority changed from Highest to Priority14

#44 - 01/25/2024 05:59 PM - Andreas Müller

- Priority changed from Priority14 to Priority13

#45 - 01/29/2024 11:47 PM - Andreas Müller

- Status changed from Resolved to Closed

- Target version deleted (Release 5.42)

- % Done changed from 50 to 100

All databases are exported