Access to Memory (AtoM) - Feature #12281

Add roundtrip option to command-line CSV import task for better matching when updating in a single system

06/19/2018 04:58 PM - Steve Breker

Status:	Verified		**Start date:**	06/19/2018
Priority:	Medium		**Due date:**	
Assignee:			**% Done:**	100%
Category:	CSV import		**Estimated time:**	16.00 hours
Target version:	Release 2.6.0			
Google Code Legacy ID:			**Tested version:**	1.3.1
Sponsored:	No		**Requires documentation:**	

Description

This ticket stems from a Groups discussion where a user attempted to export a CSV of descriptions, modified the identifier values in the CSV, and attempted to re-import the CSV back into AtoM.

https://groups.google.com/forum/#!topic/ica-atom-users/6gZjcgBl6q0

These records failed to import because:

- Keymap matching is not applicable, and will fail because legacyId in the CSV will be the AtoM information object id and not the original source_id.
- secondary matching on the combination of title, identifier and repo name will fail because the identifier was changed in the CSV.

This change will add new matching logic specifically to support CSV round-tripping:
- Add logic to lib/QubitFlatfileImport.class.php to directly match from legacyId to informationObject.id when a new CLI param is set (--roundtrip).
- This feature will be available on the CLI only.
- When --roundtrip is set, display a warning like the upgrade-sql warning indicating to the user they should only use this if it's actually a round trip, and that they have made a DB backup beforehand.
- Allow a 'force-silent' option to suppress this warning in case it needs to be scripted.
- Update the AtoM docs to include this feature and provide some detail on how matching works, and when each type of matching is used.
- When --roundtrip is set, do not attempt keymap matching or secondary title, identifier, reponame matching - only match legacyId to info obj id.
- Do not create a keymap record when --roundtrip is set.
- Audit log should reflect this record update when the CSV is loaded (e.g. "description was updated").
- Ensure CSV records that are still unmatched even when using --roundtrip are not imported.

History

#1 - 03/14/2019 02:27 PM - Dan Gillean

- *Estimated time changed from 20.00 to 24.00*

#3 - 03/29/2019 01:34 PM - Dan Gillean

- *Estimated time changed from 24.00 to 40.00*

#5 - 08/02/2019 08:54 PM - Mike Cantelon

- *Status changed from New to Code Review*

- *Assignee set to Steve Breker*

Hi Steve... assigning this to you for CR as you figured out how to do it and designed the feature.

PR for CR: https://github.com/artefactual/atom/pull/944

#6 - 08/14/2019 03:08 PM - Steve Breker

- Status changed from Code Review to Feedback

- Assignee changed from Steve Breker to Mike Cantelon

CR complete. Looks good to me!

#7 - 08/14/2019 03:20 PM - Mike Cantelon

- Status changed from Feedback to QA/Review

- Assignee deleted (Mike Cantelon)

Merged into qa/2.6.x.

#8 - 01/03/2020 11:26 AM - Dan Gillean

- Project changed from AtoM Wishlist to Access to Memory (AtoM)

- Subject changed from CSV import roundtrip feature to Add roundtrip option to command-line CSV import task for better matching when updating in a single system

- Category set to CSV import

- Target version set to Release 2.6.0

- Estimated time changed from 40.00 to 16.00

- Requires documentation set to Yes

Moving this to main AtoM project now that we've implemented it for the CLI, so we remember to test and document it. Updated the title to reflect the fact that this is not yet supported in the UI.

#9 - 03/27/2020 02:23 PM - Dan Gillean

- Status changed from QA/Review to Verified

- % Done changed from 0 to 100

- Requires documentation deleted (Yes)

- Tested version 1.3.1 added

Added to 2.6 AtoM documentation in: https://github.com/artefactual/atom-docs/commit/11ebe2a1165b95cbe9c138f599dff30dc2f432d8