Migrating Data from Archivists' Toolkit to ArchivesSpace Using the Migration Tool
last updated December 19, 2017

These guidelines are for migrating data from Archivists' Toolkit 2.0 Update 16 to all ArchivesSpace 2.1.x or 2.2.x releases using the migration tool provided by ArchivesSpace. Migrations of data from earlier versions of the Archivists' Toolkit (AT) or other versions of ArchivesSpace are not supported by these guidelines or migration tool.

Note: A migration from Archivists' Toolkit to ArchivesSpace should not be run against an active production database.

A. Preparing for migration

- Make a copy of the AT instance, including the database, to be migrated and use it as the source of the migration. It is strongly recommended that you not use your AT production instance and database as the source of the migration for the simple reason of protecting the production version from any anomalies that might occur during the migration process.

- Review your source database for the quality of the data. Look for invalid records, duplicate name and subject records, and duplicate controlled values. Irregular data will either be carried forward to the ArchivesSpace instance or, in some cases, block the migration process.

- Select a representative sample of accession, resource, and digital object records to be examined closely when the migration is completed. Make sure to represent in the sample both the simplest and most complicated or extensive records in the overall data collection.

- Note:

   - An AT subject record will be set to type 'topical' if it does not have a valid AT type statement or its type is not one of the types in ArchivesSpace. Several other AT LookupList values are not present in ArchivesSpace. These LookupList values cannot be added during the AT migration process and will therefore need to be changed in AT prior to migration. For full details on enum (controlled value list) mappings see the data map. You can use the AT Lookup List tool to change values that will not map correctly, as specified by the data map.

   - Record audit information (created by, date created, modified by, and date modified) will not migrate from AT to ArchivesSpace. ArchivesSpace will assign new audit data to each record as it is imported into ArchivesSpace. The exception to this is that the username of the user who creates an accession

record will be migrated to the accession general note field.

- Implement an ArchivesSpace production version including the setting up of a MySQL database to migrate into. Instructions are included at http://archivesspace.github.io/archivesspace/user/getting-started/ and http://archivesspace.github.io/archivesspace/user/running-archivesspace-against-mysql/.

B. Preparing for Migrating AT Data

- The migration process is iterative in nature. A migration report is generated at the end of each migration routine. The report indicates errors or issues occurring with the migration. (An example of an AT migration report is provided at the end of this document.) You should use this report to determine if any problems observed in the migration results are best remedied in the source data or in the migrated data in the ArchivesSpace instance. If you address the problems in the source data, then you can simply conduct the migration again.

- However, once you accept the migration and address problems in the migrated data, you cannot migrate the source data again without establishing a new target ArchivesSpace instance. Migrating data to a previously migrated ArchivesSpace database may result in a great many duplicate record error messages and may cause unrecoverable damage to the ArchivesSpace database.

- Please note, data migration can be a very memory and time intensive task due to the large number of records being transferred. As such, we recommend running the AT migration on a computer with at least 2GB of available memory.

- Make sure your ArchivesSpace MySQL database is setup correctly, following the documentation in the ArchivesSpace README file. When creating a MySQL database, you MUST set the default character encoding for the database to be UTF8. This is particularly important if you use a MySQL client, such as Navicat, MySQL Workbench, phpMyAdmin, etc., to create the database. See http://archivesspace.github.io/archivesspace/user/running-archivesspace-against-mysql/ for more details.

- Increase the maximum Java heap space if you are experiencing time out events. To do so:

  - Stop the current ArchivesSpace instance
  - Open in a text editor the file "archivesspace.sh" (Linux / Mac OSX) or archivesspace.bat (Windows). The file is located in the ArchivesSpace installation directory.
  - Find the text string "-Xmx512m" and change it to "-Xmx1024m".

- ○ Save the file.
- ○ Restart the ArchivesSpace instance.
- ○ Restart the AT migration process.

C. Running the Migration Tool as an AT Plugin

- ● Make sure that the AT instance you want to migrate from is shut down. Next, download the "scriptAT.zip" file from the at-migration release github page (https://github.com/archivesspace/at-migration/releases) and copy the file into the plugins folder of the AT instance, overwriting the one that's already there if needed.

- ● Make sure the ArchivesSpace instance that you are migrating into is up and running.

- ● Restart the AT instance to load the newly installed plug-in. To run the plug-in go to the "Tools" menu, then select "Script Runtime v1.0", and finally "ArchivesSpace Data Migrator". This will cause the plug-in window to display.

- Change the default information in the Migrator UI:

  - Threads -- Used to specify the number of clients that are used to copy Resource records simultaneously. The limit on the number of clients depends on the record size and allocated memory. A number from 4 to 6 is generally a good value to use, but can be reduced if an "Out of Memory Exception" occurs.
  - Host -- The URL and port number of the ArchivesSpace backend server
  - "Copy records when done" checkbox – Used to specify that the records should be copied once the repository check has completed.
  - Password -- password for the ArchivesSpace "admin" account. The default value of "admin" should work unless it was changed by the ArchivesSpace administrator.
  - Reset Password -- Each user account transferred has its password reset to this. Please note that users need to change their password when they first log-in unless LDAP is used for authentication.
  - "Specify Type of Extent Data" Radio button – If you are using the BYU Plugin, select that option. Otherwise, leave as the default – Normal or Harvard Plugin.
  - Specify Unlinked Records to NOT Copy checkboxes – If you have name or subject records that are not linked to accessions, resources, or digital objects, you can choose not to migrate those to ArchivesSpace.
  - "Records to Publish?" checkboxes – Used to specify what types of records should be published after they are migrated to ArchivesSpace.
  - Text box showing -refid_unique, -term_default – This is needed for the functioning of the migration tool. Please do not make changes to this area.
  - Output Console – Display section for following the migration while it is running
  - View Error Log -- Used to view a printout of all the errors encountered during the migration process. This can be used while the migration process is underway as well.

- Once you have made the appropriate changes to the UI, there are three buttons to choose from to start the migration process:
  1. Copy to ArchivesSpace – This starts the migration to the ArchivesSpace instance indicated by the Host URL.
  2. Run Repository Check – The repository check searches for, and attempts to fix repository misalignment between Resources and linked Accession/Digital Object records. The fix applied entails copying the linked accession/digital object record to the repository of the resource record in the ArchivesSpace database (those record positions are not modified in the AT database).

     As long as accession records are not linked to multiple Resource records in different repositories, the fix will be valid. Otherwise, you will receive a warning message. For such cases, the Resource and Accession record(s) will still be migrated, but without links to one another; those links will need to be re-established in ArchivesSpace.

This misalignment problem involves only accession and resource records and not digital object records, as accession and resource records have a many-to-many relationship. Assessments also can have a many-to-many relationship with resources, accessions, and digital objects. However, since assessments are small and quick to copy, they will simply be copied to as many repositories as needed to establish all the appropriate links.

If the "Copy Records When Done" checkbox is selected, the records will be migrated to the ArchivesSpace instance once the check is completed.

3. Continue Previous Migration – If the migration process fails, this is used to skip to the place the failed previous migration left off. This should allow the migration process of resource records to be gracefully restarted without having to clean out the ArchivesSpace backend database and start from scratch.

- For most part, the data migration process should be automatic, with an error log being generated when completed. However, depending on the particular data, various errors may occur that would require the migration to be re-run after they have been resolved by the user. The time a migration takes to complete will depend on a number of factors (database size, network performance etc.), but can be anywhere from a couple of hours to a few days.

- Data from the following AT modules will migrate:

  - Lookup Lists
  - Repositories
  - Locations
  - Users
  - Subjects
  - Names
  - Accessions
  - Digital Object and Digital Object Components
  - Resources and Resource Components
  - Assessments

- Data from the following AT modules will not migrate
  - Reports

D. Assessing the Migration and Cleaning Up Data

- Use the migration report to assess the fidelity of the migration and to determine whether to:

  - Fix data in the source AT instance and conduct the migration again.

or
- ○ Fix data in the target ArchivesSpace instance.

If you select to fix the data in AT and conduct the migration again, you will need to delete all the content in the ArchivesSpace instance.

If you accept the migration in the ArchivesSpace instance, the following outlines how to check and fix your data.

- ● Re-establish user passwords. While user records will migrate, the passwords associated with them will not. You will need to re-assign those passwords according to the policies or conventions of your repositories.

- ● Review closely the set of sample records you selected:

  - ○ Accessions
  - ○ Resources
  - ○ Digital objects

- ● Review the following groups of records, making sure the correct number of records migrated:

  - ○ Accessions
  - ○ Assessments
  - ○ Resources
  - ○ Digital objects
  - ○ Controlled vocabulary lists
  - ○ Subjects
  - ○ Agents (Name records in AT)
  - ○ Locations
  - ○ Collection Management Classifications
  - ○ There may be a few extra agent records due to ArchivesSpace defaults or extra assessments if they were linked to records from more than one repository.

- ● In conducting the reviews, look for duplicate or incomplete records, broken links, or truncated data.

- ● Take special care to check to make sure your container data and locations are correct. The model for this is significantly different between AT and ArchivesSpace (where locations are tied to a container rather than directly to a resource or accession), so this presents some challenges for migration.

- ● Merge enumeration values as necessary. For instance, if you had both 'local' and 'local sources' as a source for names, it might be a good idea to merge these values.