Guidelines for Migrating Data from Archivists' Toolkit to ArchivesSpace

These migration guidelines are for migrating data from Archivsts' Toolkit 2.0 update 15 to all ArchivesSpace 2.x releases. Migrations of data from earlier versions of the Archivists' Toolkit or to later versions of ArchivesSpace are not supported by these migration guidelines or the AT to ArchivesSpace migration plugin.

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 to be migrated and use it as the source of the migration. It is strongly recommended that you not use your AT production version as the source of the migration for the simple reason of protecting the production version from any anomalies that might occur during the migration routine.

● Make sure your 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 'Set up your MySQL Database' for more details.

● 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 routine.

For instance 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 LookupList values are not present in ArchivesSpace and can not be added and will therefore need to be changed. For full details on enum 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.

- 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.

- While the following assessments have no direct bearing on the actual migration, they may help you understand better the workload associated with migrating the data, as well as determine the optimum time for conducting the migration

  ○ How will the migration affect your current ability to generate metadata records such as EAD and MARCXML for resources and METS, MODS, Dublin Core, and MARCXML records for digital objects?
  ○ How will the migration affect your current ability to generate reports of the data?
  ○ To what extent will the migration require you to revise any stylesheets you use for rendering AT data?

- 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

B. Migrating AT Data

- The migration process may be more or less 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.

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

ArchivesSpace Instructions:

- Implement an ArchivesSpace production version (see "Running ArchivesSpace against MySQL" section of page at http://archivesspace.github.io/archivesspace/doc/ and, for the daily builds, http://aspace.hudmol.com/build-snapshots/):

Running ArchivesSpace against MySQL

- Out of the box, the ArchivesSpace distribution runs against an embedded database, but this is only suitable for demonstration purposes. When you are ready to start using ArchivesSpace with real users and data, you should switch to using MySQL. MySQL offers significantly better performance when multiple people are using the system, and will ensure that your data is kept safe.

Download MySQL Connector

- ArchivesSpace requires the MySQL Connector for Java, which must be downloaded separately because of its licensing agreement. Download the Connector and place it in a location where ArchivesSpace can find it on its classpath:

```
$ cd lib
$ curl -Oq http://repo1.maven.org/maven2/mysql/
```

```
mysql-connector-java/5.1.24/mysql- connector-java-5.1.24.jar
```

Note that the version of the MySQL connector may be different by the time you read this.

Set up your MySQL database

- Next, create an empty database in MySQL and grant access to a dedicated ArchivesSpace user (this example uses username as and password as123):

```
$ mysql -uroot -p

mysql> create database archivesspace default character set utf8;
Query OK, 1 row affected (0.08 sec)

mysql> grant all on archivesspace.* to 'as'@'localhost'
identified by 'as123';
Query OK, 0 rows affected (0.21 sec)
```

- Then, modify your config/config.rb file to refer to your MySQL database:

```
AppConfig[:db_url]="jdbc:mysql://localhost:3306/archivesspace?use
r=as&password=as123&useUnicode=true&characterEncoding=UTF-8"
```

- There is a database setup script that will create all the tables that ArchivesSpace requires. Run this with:

```
scripts/setup-database.sh
```

Or for windows:

```
setup-database.bat
```

Start ArchivesSpace

- Once your database is configured, start the application using archivesspace.sh (or archivesspace.bat under Windows).

- Confirm that ArchivesSpace is running correctly by accessing the following URLs in your browser:

```
http://localhost:8089/ – the backend
http://localhost:8080/ – the staff interface
http://localhost:8081/ – the public interface
```

`http://localhost:8090/` – the Solr admin console

Running Migration Tool as AT Plugin

- Install AT 2.0 Update 15, or above (AT Beta Download) which is required to run the plugin.

- Close out of AT and download and copy the "scriptAT.zip" file into the plugins folder of the AT, overwriting the one that's already there if needed.

- Once the zip file is in place, restart the AT to load newly installed plug-in. To run the plug-in go to the "Tools" menu, then select "Script Runtime v1.0", and finally "Archives Space Data Migrator". This will cause the plug-in window (see image below) to display.

- Enter the required information in the required input fields then press the "Copy to ArchivesSpace" button to start the migration process or the "Run repository check" button if you want to check for and correct repository mismatches first (explained below).

- 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, then 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 particular aspect of the 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.

- The purpose of the input fields is as follows:

  - "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 of 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 Archives Space backend server
  - "Run Repository check" -- Used to run the repository check as explained above.

- ○ " Copy records when done" box – Used to specify that the records should be copied once the repository check has been done.
  - ○ "Continue Previous Migration"-- Used to skip straight to the place a previous migration left off. This should allow the migration process of resource records to be gracefully restarted without having to clean out the Archives Space backend database and start from scratch.
  - ○ "Password" -- password for the Archives Space "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.
  - ○ "View Error Log" -- Used to view printout of all the errors encountered during the migration process. This can be used while the migration process is underway as well.

- For most part, the data migration process should be automated, 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 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

C. Assessing the Migration and Cleaning Data Up

● 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 record groups, 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 resource locations are correct. The model for this is significantly different between AT and ArchivesSpace (the locations are tied to a container rather than directly to a resource or accession), so this presents some challenges for migration.

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