

Guidelines for Migrating Data from Archivists' Toolkit to ArchivesSpace

These migration guidelines are for migrating data from Archivists' Toolkit 2.0 update 15 to all ArchivesSpace 1.0.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

- 1: 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.
- 2: Make sure your MySQL database is set up 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.

```
$ 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. When you modify your configuration file, **MAKE SURE** you specify that the character encoding for the database to be UTF-8 as illustrated below (the following should be all on one line):

[Type text]

[Type text]

[Type text]

```
AppConfig[:db_url] = "jdbc:mysql://localhost:3306/archivesspace?  
user=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 setup-database.bat under Windows
```

- 3: 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.
 - An AT subject record will not migrate if it does not have a valid AT type statement. You can use the AT Lookup List tool to see what subject types are not valid and to change invalid expressions to valid expressions. Any term in the AT list of subject term types that does not match one of the six terms in the screen clip below will need to be transformed to one of those six terms.

2 | [Type the document title]

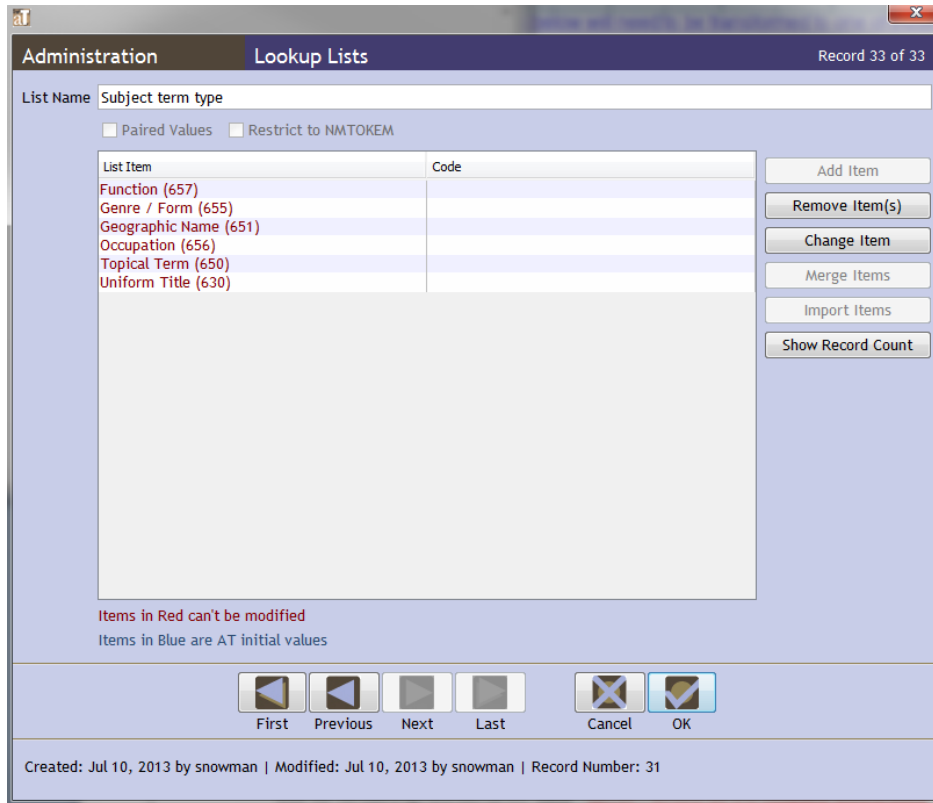


archivesspace.org

[Type text]

[Type text]

[Type text]



- 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.

4: 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.

5: 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:

3 | [Type the document title]



[Type text]

[Type text]

[Type text]

- a) 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?
- b) How will the migration affect your current ability to generate reports of the data?
- c) To what extent will the migration require you to revise any stylesheets you use for rendering AT data?

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

- a) Stop the current ArchivesSpace instance
- b) 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
- c) Find the text string “-Xmx512m” and change it to “-Xmx1024m”.
- d) Save the file
- e) 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 below.)

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.

4 | [Type the document title]



[Type text]

[Type text]

[Type text]

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.

Archivists' Toolkit Instructions:

- 1) 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):

5 | [\[Type the document title\]](#)



[Type text]

[Type text]

[Type text]

```
$ 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?  
user=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 setup-database.bat under Windows
```

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

6 | [\[Type the document title\]](#)



[Type text]

[Type text]

[Type text]

to display.

- Enter the required information in the required input fields then press the “Copy to Archives Space” button to start the migration process.

The screenshot shows a window titled "Archives Space Data Migrator Beta (v09-10-2013B)". The window contains the following elements:

- Archives Space Version: v1.0.0
- Threads: 4
- Copy To Archives Space button
- Host: http://localhost:8089
- Run Repository Check button
- Copy Records When Done
- View Report button
- Continue From Resource Records
- Reset Password
- Password: admin
- archive
- Output Console: (empty text area)
- View Error Log button
- Errors: N/A
- Cancel Copy button
- Close button

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

7 | [Type the document title]



[Type text]

[Type text]

[Type text]

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 search for, and attempt 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 an error will still occur, as the fix cannot be scripted. 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 the ArchivesSpace the link to the particular accession record. 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. "
- “Copy records when done” box – Used to specify that the records should be copied once the repository check has been done.
- “Continue From Resource Records”-- When selected, the program skips straight to migrating Resource records, provided all other records have already been successfully transferred to the Archives Space backend. 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.

8 | [Type the document title]



[Type text]

[Type text]

[Type text]

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

Data from the following AT modules will not migrate

- Assessments
- Reports

C: Assessing the Migration and Cleaning Data Up

- 1: Use the migration report to assess the fidelity of the migration and to determine whether to
 - 1A: fix data in the source AT instance and conduct the migration again, or
 - 1B: fix data in the target ArchivesSpace instance.

If you select no. 1, you will need to delete all the content in the ArchivesSpace instance.

If you accept the migration in the ArchivesSpace instance, then do the following:

- 2: 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.
- 3: Review closely the set of sample records you selected:

9 | [\[Type the document title\]](#)



[Type text]

[Type text]

[Type text]

- Accessions
- Resources
- Digital objects

4: Review the following record groups, making sure the correct number of records migrated:

- Accessions
- Resources
- Digital objects
- Controlled vocabulary lists
- Subjects
- Agents (Name records in AT)
- Locations
- Collection Management
- Classifications

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



REPORT EXAMPLE: Migrating AT data to Archives Space

999 RECORD CONVERSION ISSUES

ACCESSIONS

ID Cleaned Up (Empty spaces removed)

Acc1.345.678.789

Acc2.345.678.789

...

ID DE-Duped (Extending base id with random characters to make unique.)

Acc1.345.678.789 -> Acc1 ##dfg.345.678.789

Acc2.345.678.789 -> Acc1 ##fvc.345.678.789

...

DIGITAL OBJECTS

Missing Id Added (Added ID; Title from incoming DO record)

Digital Object ID ##cvgcvb; Start of NYC St. Patrick's Day Parade

ID DE-Duped (Extending id with random characters to make unique.)

DO_Ms20 -> DO_Ms20 ##fds

DO_Ms21 -> DO_Ms21 ##sxe

...

RESOURCES

ID Cleaned Up (Empty spaces removed)

Res1.345.678.789

Res2.345.678.789

...

ID DE-Duped (Extending base id with random characters to make unique.)

Res1.345.678.789 -> Res1 ##dfg.345.678.789

Res2.345.678.789 -> Res1 ##fvc.345.678.789

...

EAD_ID DE-Duped (Extending id with random characters to make unique.)

FA_Ms20 -> FA_Ms20 ##fds

FA_Ms21.xml -> FA_Ms21.xml ##sxe

...

REPOSITORY MISALIGNMENT (A resource record in one repository is linked to an accession or digital object record in a different repository)

Resource.id.3.4 [Huntington Library] :: Digital.Object.id.1 [AT]

Resource.id.3.4 [Huntington Library] :: Accession.id.1 [AT]

LINKED RECORDS NOT MIGRATED (Unable to create link)

[Type text]

[Type text]

[Type text]

```
Accession.Id.1 :: Subject ABC
Accession.id.1 :: Name ABC
Digital.Object.id.1 :: Subject ABC
Digital.Object.id.1 :: Name ABC
Resource.id.3.4 :: Accession.id.1
Resource.id.3.4 :: Digital.Object.id.1
Resource.id.3.4 :: Subject ABC
Resource.id.3.4 :: Name ABC
```

32 RECORDS NOT MIGRATED

```
Endpoint: http://localhost:8089/users?password=admin&groups%5B%5D=
AT Identifier:N/A
Status code: 400
Status text: Bad Request
```

```
{"error":{"username":["Username 'asadmin' is already in use"]}}
```

```
Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->Slavery in Alabama
Status code: 400
Status text: Bad Request
```

```
{"error":{"terms":["Subject must be unique"]}}
```

```
Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->United States--History--Revolution, 1775-1783--Fiction
Status code: 400
Status text: Bad Request
```

```
{"error":{"terms":["Subject must be unique"]}}
```

```
Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->United States--History--Revolution, 1775-1783--
Participation, German--Fiction
Status code: 400
Status text: Bad Request
```

```
{"error":{"terms":["Subject must be unique"]}}
```

```
Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->Women's Shoes
Status code: 400
Status text: Bad Request
```

```
{"error":{"source":["Invalid value 'localb'. Must be one of: aat, aat.,
abs_rubriek, abs_trefwoord, dictionary of occupational titles, georef thesaurus,
gmGPC, ingest, itoamc, lcna, lcnaf, lcsh, lctgm, local, local broad, mesh, nssha,
nwda, rbgenr, rbtyp, riamco, source not specified, tgn, umabroad, unspecified, vmf,
other_unmapped"]}, "warning": {}, "invalid_object": "#<JSONModel(:subject)
{\\\"source\\\"=>\\\"localb\\\", \\\"vocabulary\\\"=>\\\"/vocabularies/1\\\", \\\"external_ids\\\"=>[{\\"}
```

12 | [Type the document title]

[Type text]

[Type text]

[Type text]

```
"external_id"=>"2030", "source"=>"Archivists Toolkit Database::SUBJECT"}, {"terms"=>[{"vocabulary"=>"/vocabularies/1", "term"=>"Women's Shoes", "term_type"=>"topical"}], "jsonmodel_type"=>"subject", "external_documents"=>[]}>"}
```

Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->TEst
Status code: 400
Status text: Bad Request

```
{"error":{"source":["Invalid value 'localb'. Must be one of: aat, aat., abs_rubriek, abs_trefwoord, dictionary of occupational titles, georef thesaurus, gmgpc, ingest, itoamc, lcna, lcnaf, lcsh, lctgm, local, local broad, mesh, nssha, nwda, rbgenr, rbtyp, riamco, source not specified, tgn, umabroad, unspecified, vmf, other_unmapped"]},"warning":{},"invalid_object":"#<JSONModel(:subject) {"source"=>"localb", "vocabulary"=>"/vocabularies/1", "external_ids"=>[{"external_id"=>"2260", "source"=>"Archivists Toolkit Database::SUBJECT"}, {"terms"=>[{"vocabulary"=>"/vocabularies/1", "term"=>"TEst", "term_type"=>"other_unmapped"}], "jsonmodel_type"=>"subject", "external_documents"=>[]}>"}
```

Endpoint: http://localhost:8089/subjects
AT Identifier:Subject->Korean War, 1950-1953
Status code: 400
Status text: Bad Request

```
{"error":{"terms":["Subject must be unique"]}}
```

Endpoint: http://localhost:8089/repositories/78/accessions
AT Identifier:Accession->68/2010
Status code: 400
Status text: Bad Request

```
{"error":{"deaccessions/1/description":["Must be 255 characters or fewer"]},"warning":{},"invalid_object":"#<JSONModel(:accession) {"use_restrictions_note"=>"The collection is available for research.", "restrictions_apply"=>false, "external_ids"=>[{"external_id"=>"6949", "source"=>"Archivists Toolkit Database::ACCESSION"}, {"collection_management"=>{"processing_priority"=>"medium", "processors"=>"Michelle Brundige", "rights_determined"=>true, "processing_status"=>"other_unmapped", "external_ids"=>[]}, {"deaccessions"=>[{"extents"=>[{"extent_type"=>"photographs", "container_summary"=>"All photographs were transferred to the Photo Archives department.", "number"=>"50.0", "portion"=>"whole"}], "scope"=>"whole", "reason"=>"Photographs are to be stored in the Photo Archives Department at The Ohio State University Archives per the department's policy.", "description"=>"All photographs were transferred to the Photo Archives department.", "notification"=>false, "date"=>{"expression"=>"2012-02-28", "era"=>"ce", "label"=>"deaccession", "date_type"=>"single", "beg
```

13 | [Type the document title]

[Type text]

[Type text]

[Type text]

```
in\"=>\"2012-02-28\"}},
{\"extents\"=>[{\\"extent_type\"=>\"linear_foot\", \\"container_summary\"=>\"During
the processing of the collection revealed some materials determined to be of a
personal nature, such as tax documents, insurance documents and contracts. These
items were returned to Donald Harris. All documents relating to Donald Harris%Ūa
career at the New England Conservatory of Music and the University Of Hartford
Hartt School Of Music were also
returned.\\n\", \\"number\"=>\"2.0\", \\"portion\"=>\"whole\"}], \\"scope\"=>\"whole\"
, \\"reason\"=>\"Per The Ohio State University Archives Statement of Authority,
records not pertaining to Donald Harris' career at The Ohio State University are
not kept. An exception was made for materials relating to Donald Harris' book on
Berg and Schoenberg.\\n\", \\"description\"=>\"During the processing of the collection
revealed some materials determined to be of a personal nature, such as tax
documents, insurance documents and contracts. These items were returned to Donald
Harris. All documents relating to Donald Harris%Ūa career at the New England
Conservatory of Music and the University Of Hartford Hartt School Of Music were
also returned.\\n\", \\"notification\"=>true, \\"date\"=>{\\"expression\"=>\"2010-08-
05\", \\"era\"=>\"ce\", \\"label\"=>\"deaccession\", \\"date_type\"=>\"single\", \\"beg
in\"=>\"2010-08-
05\"}}}], \\"extents\"=>[{\\"extent_type\"=>\"other_unmapped\", \\"number\"=>\"0\", \\"p
ortion\"=>\"whole\"}], \\"suppressed\"=>false, \\"title\"=>\"Donald Harris' Faculty
Papers\", \\"resource_type\"=>\"papers\", \\"acquisition_type\"=>\"gift\", \\"id_0\"=>
\"68/2010\", \\"linked_agents\"=>[{\\"ref\"=>\"/agents/people/667\", \\"role\"=>\"subj
ect\", \\"terms\"=>[]}],
{\\"ref\"=>\"/agents/people/667\", \\"role\"=>\"creator\", \\"terms\"=>[]}], \\"accessi
on_date\"=>\"2010-07-
14\", \\"access_restrictions\"=>true, \\"jsonmodel_type\"=>\"accession\", \\"publish\"
=>false, \\"subjects\"=>[], \\"dates\"=>[], \\"external_documents\"=>[], \\"rights_stat
ements\"=>[], \\"related_resources\"=>[], \\"use_restrictions\"=>false, \\"instances\"
=>[]>\"}
```

Endpoint: http://localhost:8089/repositories/4/accessions
AT Identifier:Accession->z.2013.004.BC
Status code: 400
Status text: Bad Request

```
{\"error\":{\"id_0\":[\"That ID is already in use\"]}}
```

Finish copying data ... Total time: 3 hr 39 min 08.91 sec

Number of Records copied:
Lookup List : 16 / 34 (47.06%)
Repositories : 127 / 127 (100.00%)
Locations : 2835 / 2835 (100.00%)
Users : 90 / 90 (100.00%)
Subjects : 1943 / 1949 (99.69%)
Names : 1429 / 1429 (100.00%)
Accessions : 1725 / 1727 (99.88%)
Digital Objects : 151 / 157 (96.18%)
Resource Records : 690 / 707 (97.60%)

14 | [Type the document title]