Guidelines for Migrating Data from Archon to ArchivesSpace

[Preparing for migration](#h.dkb03bxmep7)

[Migrating Archon Data](#h.blqyi2h8upq5)

[Install Archon update](#h.ffmcvhbofbza)

[Install Archon to ASpace Migration Tool](#h.erdgugbb7l2o)

[Install ArchivesSpace Instance](#h.3aznzagss61y)

[Prepare to Launch Migration](#h.cgxof5xvigt2)

[Launch Migration Process](#h.2w9igh7smuah)

[Assess the Migration and Clean Up Data](#h.b05oj9kztcvw)

[Appendix A: Migration Log Review and Sample](#h.50mgsck6ndv4)

These migration guidelines are for migrating data from Archon 3.21-rev2.x to ArchivesSpace 1.4.2. These guidelines and their associated Archon to ArchivesSpace migration tool do not support migrations from earlier versions of Archon, and the tools have not been tested for migration to later versions of ArchivesSpace.

# Preparing for migration

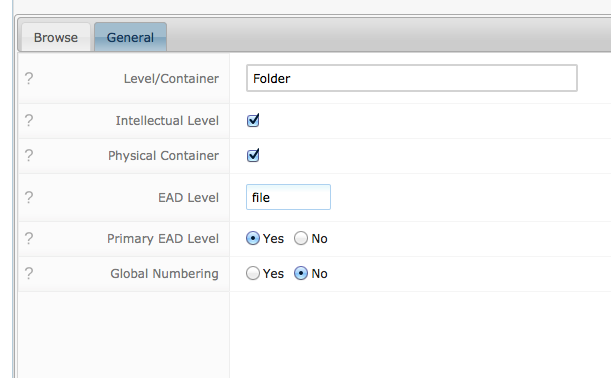
1. Review your Archon database for data quality:
2. **Accession Records**:
   1. Supply an accession date for all records, when possible. If an accession date is not recorded in Archon, the date of 01/01/9999 will be supplied during the migration process. If you wish to change this default value, you may do so by editing the following file in the new Archon distribution, prior to running the migration:

packages/core/templates/default/accession-list.inc.php

* 1. Supply an identifier for all records, when possible. If an identifier is not recorded in Archon, a supplied identifier will be constructed during the migration process, consisting of the date and the truncated accession title.

1. **Classification Records**: Ensure that there are no duplicate classification titles at the same level in the classification hierarchy. If the migration tool encounters a duplicate value, some of the save operations for classifications will fail, and you will need to redo the migration.
2. **Collection Records**: If normalized dates are not recorded correctly (i.e. if the end date and begin date are reversed), they will not be migrated or may cause the migration to fail. To check for such entries, a system administrator can run the follow query against the database:

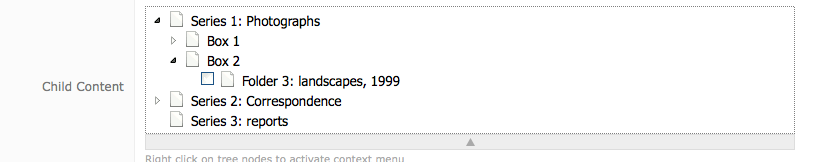
SELECT ID, Title, NormalDateBegin, NormalDateEnd FROM tblCollections\_Collections WHERE NormalDateBegin > NormalDateEnd;

1. **Level/Container Manager**: Review the settings to make sure that each ‘level container’ is appropriately marked with the correct values for “Intellectual Level” and “Physical Container” and that EAD Values are correctly recorded. 

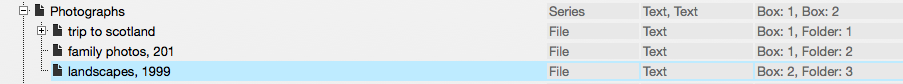
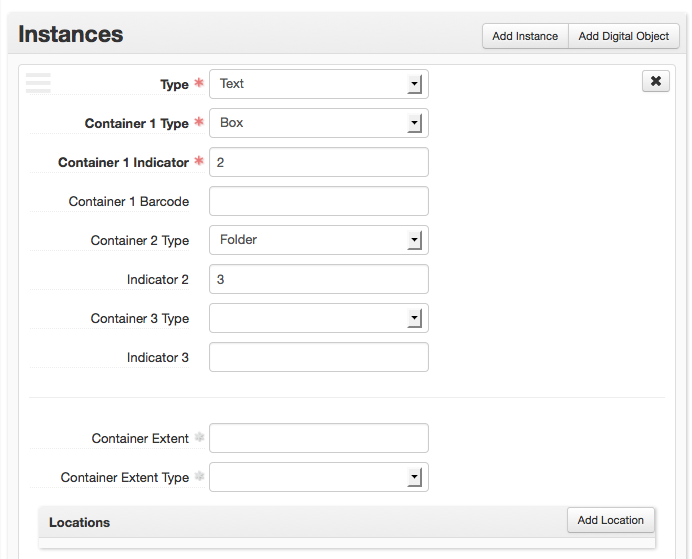
Failure to code level container values correctly may result in incorrect nesting of resource components in ArchivesSpace. While the following information does not need to be acted upon prior to migration, please note the following if you find that content is not nested correctly after you migrate:

1. Collection content records that have a level container that is ‘Intellectual Only’ will be migrated to ASpace as resource components. Each level/container that has ‘intellectual level’ checked should have a valid value recorded in the “EAD Level” field (i.e. class, collection, file, fonds, item, otherlevel, recordgrp, series, subfonds, subgrp, subseries). These values are case sensitive, and all other values will be migrated as “otherlevel” on the collection content/resource component records to which they apply.
2. Collection content records that have a level container this is ‘Physical Only’ will be migrated to ASpace as instance records of the type ‘text’ attached to a container. In ASpace, these instance/container records will be attached to the intellectual level or levels that are immediate children of container record as it was previously expressed in Archon. If the instance/container has no children it will be attached to its parent intellectual level in stead. For illustrative purposes, the following screenshots show a container record prior to and following migration.

**Box 2, Folder 3 (landscapes) in Archon:**

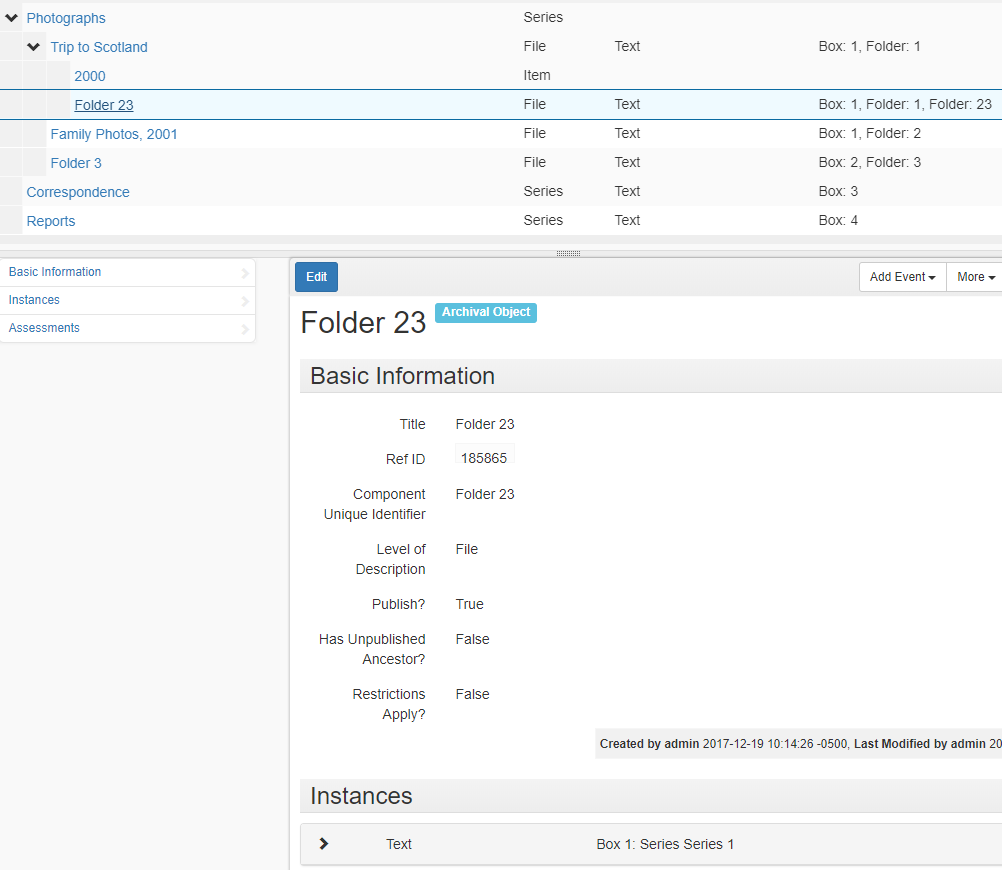
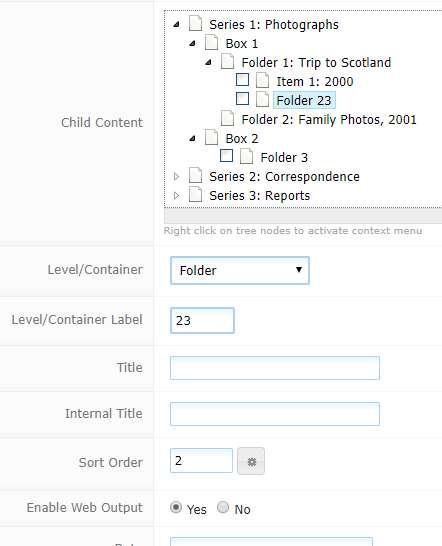


**Box 2, Folder 3 (landscapes) in ASpace with instance/container:**

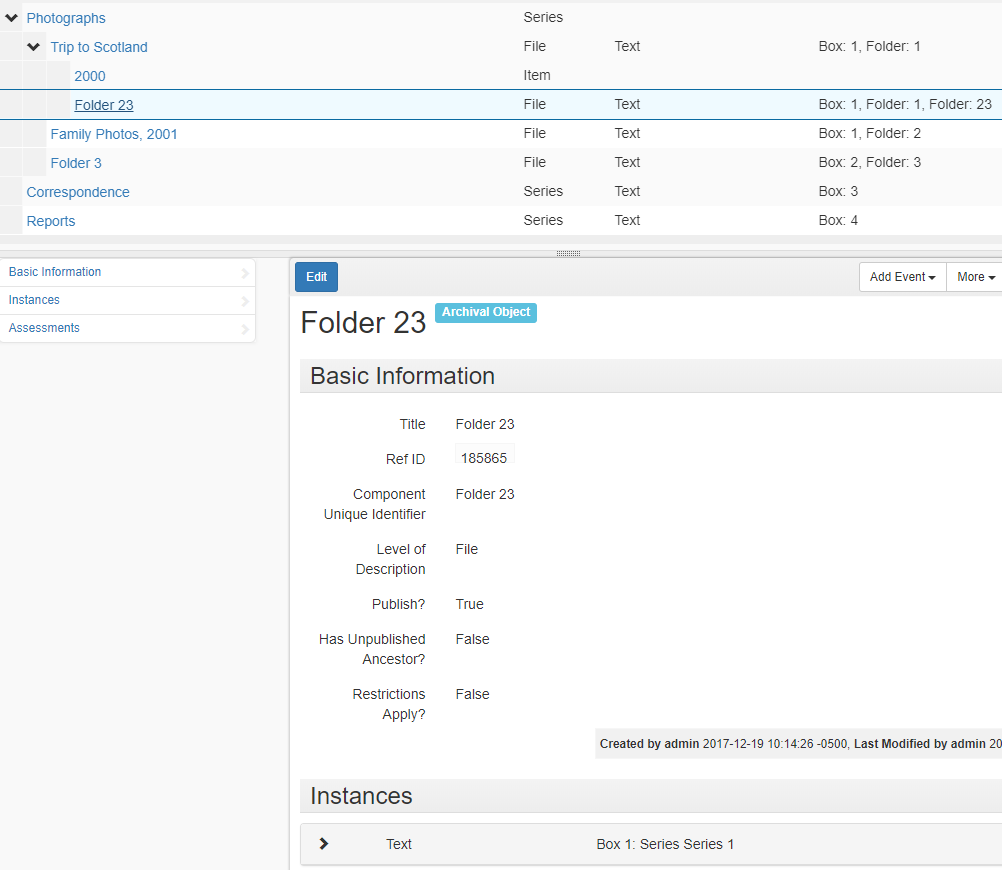
****

1. Collection content records that have both physical and intellectual levels will be migrated as both resource components and instances. In this case the instance will be attached to the resource component.
2. Collection content records that are neither physical nor intellectual levels will be migrated as if they were ‘Intellectual Only’. This is not recommended and should be fixed prior to migration.

1. **Collection Content Records**:
   1. If a value has not been set in “Title” or “Inclusive Dates” field of an “intellectual” level/container in Archon, the collection content record being migrated will be supplied a title, based on its “label” value and the “level/container” type set in Archon.



Archon Collection Content Record ASpace Resource Component Record with Instance



ASpace Collection Organization

**Untitled/Undated Content Record in Archon vs ASpace**

1. **Optional--**you may wish to complete this step if a migration fails: Check for collection content records that reference invalid ‘level/containers’. These records are found in the database tables, but are not visible to staff or end users and must be eliminated prior to migration. If not eliminated, the migration will fail. In order to identify these records, you should follow these steps. **Be very careful. If you are uncertain what you are doing, backup the database first or speak with a systems administrator!**
   1. In MySQL or SQL Server, open the table titled ‘tblCollections\_LevelContainers’. Note the ‘ID’ value recorded of each row (i.e. LevelContainer).
   2. Run a query against tblCollections\_Content to find records where the LevelID column references an invalid value. For example, if tblCollections\_Level Containers holds ‘ID’ values1-6 and 8-22:

SELECT \* FROM tblCollections\_Content Where LevelContainerID > 22 or (LevelContainerID > 6 and LevelContainerID < 8);

This will provide a list of all records with invalid levelID’ (i.e. where a record with the primary key referenced by a foreign key cannot be found).

Review this list carefully to make sure you are comfortable deleting the records, or change the LevelID to a valid integer if you wish to retain the records. If you choose to delete the records, you will need to do so directly in the database (see below.) If you choose to do the latter, you may need to take additional steps directly in the database to link these records to a valid parent content record or collection; additional instructions can be supplied upon request.

1. Run a query to delete the invalid records from the collections content table. For example:

Delete from tblCollections\_Content Where LevelContainerID > 22 or (LevelContainerID > 6 and LevelContainerID < 8);

1. **Optional**: You may wish to complete this step if the migration fails: check for ‘duplicate’ collection content records. ‘Duplicate’ records are those that occupy the same node in the collection/content hiearchy. To check for these records, run the following query in mysql or sql server.

SELECT ParentID, SortOrder, count (\*) from tblCollections\_Content Group by ParentID, SortOrder HAVING count(\*) > 1;

The query above checks for records that occupy the same branch and same position in the content hierarchy. If you discover such records, they sort order value of one of the records must be changed, so that both records occupy a unique position. In order to do this, run a query that finds all records attached to the parent record, then run an update query to change the sort order of one of the offending records so that each has a unique sort order. For example if the query above returns ParentID as a ‘duplicate’ value, you would run query one query similar to identify the offending records, and query two in to fix the problem:

**Query ONE:**

Select ID, ParentID, SortOrder, Title FROM tblCollections\_Content where ParentID=8619;

|  |  |  |  |
| --- | --- | --- | --- |
| ID | ParentID | SortOrder | Title |
| 8620 | 8619 | 1 | to mother |
| 8621 | 8619 | 1 | from mother |
| 8622 | 8619 | 3 | to father |
| 6823 | 8619 | 4 | from father |

**Query two:**

UPDATE tblCollections\_Content SET SortOrder=2 WHERE ID=8621;

1. Examine Data and Assess Potential Migration Steps
   1. Select a representative sample of accession, classification, collection, collection content, and digital object records to be examined closely when the migration is completed. Make sure to include both simple and more complicated or extensive records in the sample.
   2. While the following assessments have no direct bearing on the actual migration, they may help you understand better the workload associated migrating the data as well as determine the optimum time for conducting the migration:
   3. Assess how the migration will affect your end user views/public access to the data. If you plan to use the public access interface in ASpace, you should carefully review the default public viess supplied with ASpace. If you wish to make changes to the ASpace views, you will need to do so before migrating. Information regarding customization can be viewed at <https://github.com/archivesspace/archivesspace/tree/master/plugins>.
   4. Assess how the migration will affect your current ability to generate metadata records such as EAD and MARC for collections.
   5. Assess how the migration will affect your workflow and develop a strategy for revising workflows and documentation.

# Migrating Archon Data

The migration process may be more or less iterative in nature. In other words, you should plan to do several test migrations, culminating in a final migration. Typically, migration will require assistance from a system administrator.

A migration report is generated at the end of each migration routine and can be downloaded from the application. The report indicates errors or issues occurring with the migration. Sample data from migration report is provided in Appendix A.

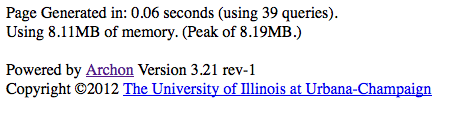
You should use this report to determine if any problems observed in the migration results are best remedied in the source data or in the migrated data in the ArchivesSpace instance. If you address the problems in the source data, then you can simply clear the database and conduct the migration again. However, once you accept the migration and make changes to the migrated data in ArchivesSpace, you cannot migrate the source data again without either overwriting the previous migration or establishing a new target ArchivesSpace instance.

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 Archon migration tool on a server with at least 2GB of available memory. Test migrations have run from under an hour to twelve hours or more in the case of complex and large instances of Archon.

## Install Archon update

**Archon 3.21 rev-2.x** includes code that produces records in a format expected by the migration tool (JSON). The migration tool described below will connect to Archon 3.21 rev-2, will read data from defined ‘endpoints,’ and will place the information in a target Archives Space instance.

1. Ensure that your current Archon installation is up to date: i.e. that you are using version 3.21 or 3.21 rev-1.
2. If yes: proceed to step d below.
3. If no:
   1. Make a copy of the source database and upgrade from your current version to 3.21, ***using the copy of the database as the source for the upgrade***. Follow the installation instructions found at <http://www.archon.org/installation.php>. If you are not running version 3.21 as your production instance, it is imperative that you work with a copy of the data when migrating forward to ArchivesSpace!!
4. Download Archon maintenance update 3.21 rev-2.x from <https://github.com/archonproject/archon/releases>.
   1. Unzip the downloaded files.
   2. Put the application files on a web server which can access the source mysql or mssql database server. Do not overwrite your production webfiles! Under normal circumstances, 3.21 rev-2 should be installed alongside an instance of 3.21 or 3.21 rev-1. The migration must use a copy of the database as its data source; you should not try to run the migration tool against a production instance of the database as this may cause data integrity issues if records are updated while the migration is ongoing..
   3. Copy config.inc.php from the root of your production instance to the root of your migration instance (you can overwrite the existing file from the distribution) and change the credentials so the you are migrating against the copy of the database. Also, turn off caching and google analytics, if you use these settings in your production instance. Save the changes back file.
5. Navigate to the URL for your new instance or Archon and login as you normally would, to test whether the system is working correctly. If the source url shows a page that looks like the screenshot below, simply append the highlighted text onto the end of the URL and reload the page: [http://example.com/archon/?settheme=default](http://example.com/archon/?settheme=default*)



## Get Archon to ASpace Migration Tool

* 1. Download the latest JAR file release from <https://github.com/archivesspace/ArchonMigrator/releases/latest>
  2. This is an executable JAR file so all you have to do is double click to run it.

## Install ArchivesSpace Instance

* 1. Download ArchivesSpace 2.2.2 (any 2.1.x or 2.2.x version should also work) from <https://github.com/archivesspace/archivesspace/releases/tag/v2.2.2>
  2. Implement an ArchivesSpace production version (see “Running ArchivesSpace against MySQL” section of page at   
     <https://github.com/archivesspace/archivesspace>). Prior to setting up the instance, please review the section on port configuration, and decide which port you would like the public and staff interface to run under.
  3. Start Archives space. Information on starting/stopping the application and running it as a Unix daemon are included in the ArchivesSpace documentation.

## Prepare to Launch Migration

**Important Note:** The migration process should be launched from a networked computer with a stable (i.e. wired) connection, and you should turn power save settings off on the client computer you use to launch the migration. So that the migration can proceed in an undisturbed fashion, you should not try to access the ArchivesSpace or Archon front end or public interface until after the migration as completed. **If you fail to follow these instructions, the migration tool may not provide useful feedback and it will be difficult to determine how successful the migration was.**

For most part, the data migration process should be automated, with errors being provided as the tool migrates and a log being made available when migration is complete. Depending on the particular data being migrated, various errors may occur These may require the migration to be re-run after they have been resolved by the user.  When this occurs, the mysql database should be emptied by the system administrator, and the migration rerun after steps are taken to resolve the problem that caused the error. Instructions on emptying the database are found in the Achivesspace documentation, but can be briefly summarized as follows:

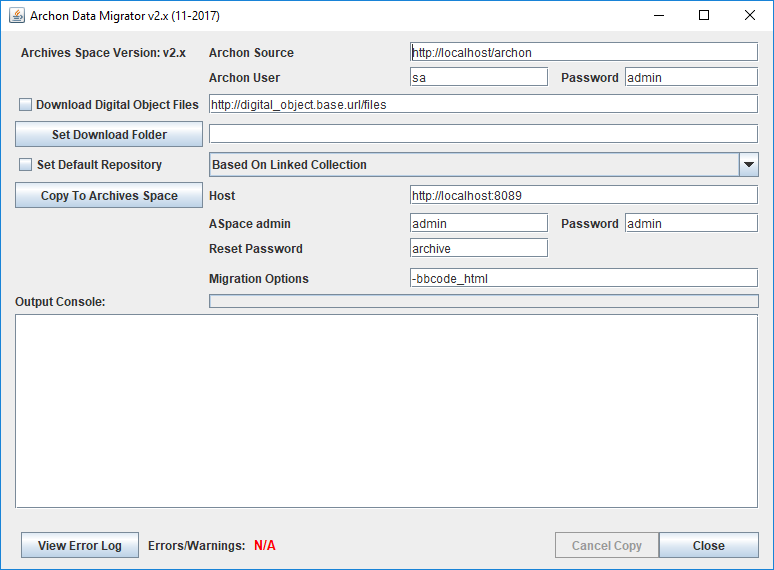
## Drop all tables from the source database using the mysql command line tools or the client application of your choice.

## Run the database installer script by running the following command from the terminal, while in the main archivesspace folder

scripts/setup-database.sh # or setup-database.bat for Windows

The time that the migration takes to complete will depend on a number of factors (database size, network performance etc.), but has been known to take anywhere from a half hour to ten or twelve hours. Most of this time will probably be spent migrating collection records.

The following Archon datatypes will migrate, and all relationships that exist between these datatypes should be preserved in ArchivesSpace, except as noted in bold below. For each datatype, post-migration cleanup recommendations are provided in parentheses:

* Editable controlled value lists:
  + Subject sources (review post migration and merge values with ASpace defaults or functionally duplicate values, when possible)
  + Creator sources (review post migration and merge values with ASpace defaults or functionally duplicate values, when possible)
  + Extent units/types (merge functionally duplicate values)
  + Material Types
  + Container Types
  + File Types
  + Processing Priorities
* Repositories
* User/logins (users will need to reset password)
* Subjects(**subjects of type personal corporate or family name are migrated as Agent record, and are linked to resources and digital objects in the subject role.** Review these records and merge with duplicate agent names from creator migration, when possible.)
* Creators/Names
* Accessions (The migration tool will supply accession identifiers, when these are blank in Archon. Review and change values, if appropriate.)
* Digital Objects: The migration tool will generate digital object metadata records in ArchivesSpace for each digital library record that is stored in your archon instance. For each file that is attached a digital library record, the migration tool will generate a digital object component and file instance record. In addition, the migration tool will provide a folder containing the source file you uploaded to archon when you created the record. In order to link these files to the digital file records in Aspace, you should place the files in a single directory on a webserver. **To preserve the linkage between the file’s metadata in archives space, you must provide the base url to the folder where the objects will be placed, for example:** 

The migration tool prepends this url to the filename to form a complete path to the object location, for each file being exported, as shown in the screenshot below. (In version 1.4.2 of ArchivesSpace, with the default digital object templates, these files will be available in the public interface by clicking a link.)

* Locations (Controlled location records are much more granular in ArchivesSpace than in Archon. You should have a location record for each unique combination of location drop down, range, section, and shelf in archon, and these records should be linked to top container records which are in turn linked to an instance for each collection where they apply.)
* Resources and Resource Components (see locations, above).

Data from the following Archon modules will not migrate to ArchivesSpace

* Books (Book data could be migrated later if a plug in is developed to support this data).
* AVSAP/Assessments

## Launch Migration Process

When you are ready to migrate, follow these instructions:

1. Enter the required information in the required input fields The purpose of the input fields are as follows:
   1. ArchonURL -- Supply the base url for the archon instance you created in step I above.
   2. Archon User: Username for an account with full administrator privileges
   3. Archon Password: Password for that same account.
   4. ASpace URL: The URL and port number of the Archives Space backend server
   5. ASpace User: username for a user account with full administrator privileges in ASpace.
   6. ASpace Password: password for the Archives Space administrator account.  The default value of “admin” should work unless it was changed by the ArchivesSpace administrator.
   7. Digital Object Base URL: Webpath to a web accessible folder where you intend to place the digital objects after the migration is complete. (see above)
   8. Set Download Folder: Specify the folder you want digital files from Archon to be downloaded to.
2. Press the “Start Migration button to start the migration process.
3. If the migration process fails: Review the error message provided and /or the migration log. Fix any issues that have been identified, clear the target MySQL database as described earlier in this document, and try again.
4. When the process has completed, download the migration.
5. Move digital objects into the folder location corresponding to the URL you provided to the migration tool as ‘Digital Object Base URL’

# Assess the Migration and Clean Up Data

1. Use the migration report to assess the fidelity of the migration and to determine whether to fix data in the source Archon instance and conduct the migration again, or fix data in the target ArchivesSpace instance. If you select to fix data in Archon, you will need to delete all the content in the ArchivesSpace instance, then rerun the migration after clearing the ArchivesSpace database.
2. Review the following record types, making sure the correct number of records migrated. In conducting the reviews, look for duplicate or incomplete records, broken links, or truncated data.
   1. Controlled vocabulary lists
   2. Classifications
   3. Accessions
   4. Resources
   5. Digital objects
   6. Subjects (not persons, families, and corporate bodies)
   7. Creators (known as Agents in ArchivesSpace)
   8. Locations
3. Review closely the set of sample records you selected, comparing data in Archon to data in ArchivesSpace.
4. If you accept the migration in the ArchivesSpace instance, then proceed to re-establish user passwords. While user records will migrate, the passwords associated with them will not. You will need to reassign those passwords according to the policies or conventions of your repositories.
5. Install and tailor public access themes and templates, if desired.

# Appendix A: Migration Log Review

The migration log provides a description of any irregularities that take place during a migration and should be saved in a secure location, for future reference. The log contains both save errors and warnings. The warnings should be reviewed after the migration for information, for potential action.

Most warnings will not require a follow up action. For example, they may note that a supplied value has been provided to meet an ArchivesSpace data model requirement. This occurs for all collections with empty identifiers. Occasionally, warnings will indicate that there was a problem establishing a link between two records for a reason such as a resource component not being found. Warnings like this should be cause for review since they may indicate that some data was lost.

Save errors will note that a particular piece of data could not be migrated because it is not supported in the ArchivesSpace data model or for some other reason. In these cases, you should review the record in Archon and in ArchivesSpace if it was migrated at all. Oftentimes, these occur due to duplicate records (such as if you have a matching creator and person subject). If a save error occurs due to a duplicate record, this is usually okay but should still be reviewed to make sure there was no data loss. If a save error occurs for any other reason, this typically means the migration will need to be rerun (unless the record it occurred on is not needed or is easier just to migrate manually).

Typically, the migration log will record the archon internal ID’s of the original Archon object being migrated whenever a save error or warning occurs. This simplifies finding and correcting relevant records.