



Database Migration Project
Import Procedures

Author: Ron Stevenhaagen

Revised by: Jeremy Heil
Alex Fletcher

Mar. 2011
Revised 13 Feb. 2014

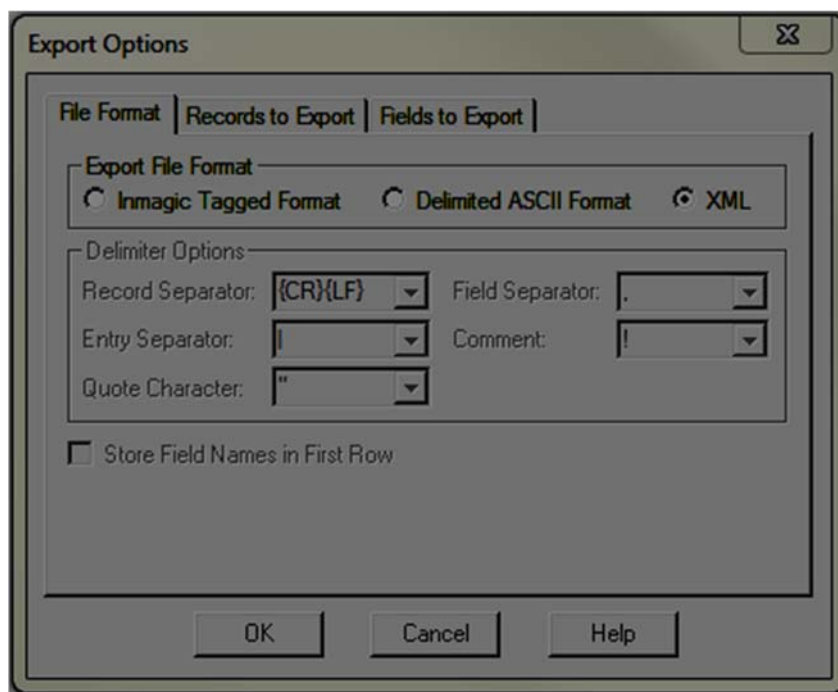
1	Data Import Process for AtOM.....	3
1.1	Prepare files from DBText.....	3
1.2	Replace with Empty Database.....	5
1.3	Confirming all files are linked from the Inventories share.....	5
1.4	Import EAC and EAD Records.....	5
2	Post-import Procedures.....	6
3	Removing Untitled Creators/Accumulators.....	7
4	Search Indexing.....	7
5	Publishing Archival Materials.....	8

1 Data Import Process for AtOM

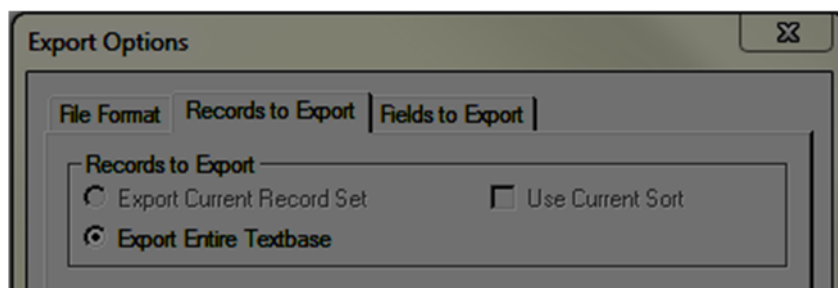
1.1 Prepare files from DBText

In order to import data into AtOM the following steps are completed in the specified order

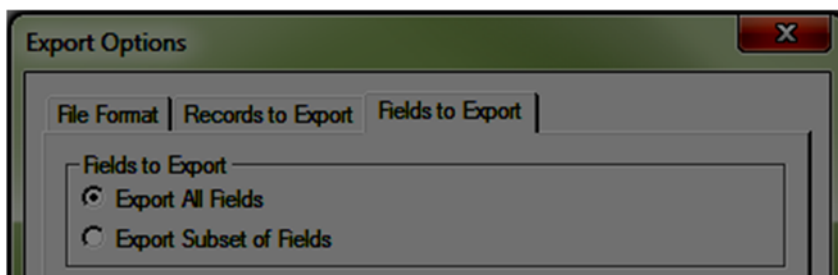
- 1) Export the following databases from InMagic using the parameters below:
 - a) Export in XML format



- b) Export entire textbase



- c) Export all fields



- d) Save files as:
 - i. Fonds, Collections and Discrete Items - save as fonds.dmp
 - ii. Box/File List - save as boxFile.dmp

2) Ensure the Inventories share is mounted by following the instructions in AtoM Installation Guide, Section 2

- 3) Copy the exported databases to the AtoM server (Ubuntu Linux), either
- using Filezilla
 - to the Inventories folder

The files should be copied to /home/stevenr/

- 4) Remove the newline characters (^M) which appear at the end of each line from the exported files by opening each file in the vi editor and issuing the following command:

```
:%s/^V^M//g
```

NOTE: ^V and ^M are CTRL-V, CTRL-M (will look like `:%s/^M//g`)

Also remove the header information from each exported file which includes all lines that appear before the line containing the following entry:

```
<inm:Record setEntry="0">
```

Once those adjustments have been made to the exported files save the file before exiting from vi.

- 5) In order to improve the performance for processing the Box File List it is necessary to break the Box File List into smaller files (one for each fonds/collection).

First remove all files in /home/stevenr/ that start with recordid_ followed by a unique five digit number by issuing the following command:

```
rm recordid_*
```

In order to create the smaller files representing each fonds/collection issue the following command from /home/stevenr/

```
perl split.pl boxFile.dmp
```

Note: the first argument to perl represents the filename of the BoxFile List dump.

1.2 Replace with Empty Database

Before the import process, the AtoM database must be empty. To do this, overwrite the existing database with atom2_users.sql

```
sudo mysql --user=root --password=[password] atom < atom2_users.sql
```

1.3 Confirming all files are linked from the Inventories share

Two reports will need to be produced to confirm that a link is established from the finding aids to those files contained in the Inventories shared directory.

In order to produce those reports run the following perl scripts

```
perl orphanedfiles.pl <fonds> <linkedrpt> <unlinkedrpt>
```

This script requires 3 arguments specifying the fonds database dump, the file name for the linked Files Report, and a file name for the unlinked Files Report.

The purpose of this Report is to show any files in the Inventories share which do not appear as a finding aid in the fonds database

```
-----  
perl unlinkedFindingAids.pl <fonds> <linkedrpt> <unlinkedrpt>
```

This script requires 3 arguments specifying the fonds database dump, the file name for the linked Files Report, and a file name for the unlinked Files Report.

The purpose of this Report is to show any finding aids which cannot be found in the Inventories share.

```
-----  
NOTE: The two reports are similar and may point to an issue with the same  
      unresolved links from both reports where there is not an exact match between the  
      finding aid and filename located in the directory. However both reports are required for  
      completeness due to the following situations:
```

- a finding aid is created where no file exists
- a file exists and no finding aid is created for this file.

1.4 Import EAC and EAD Records

- 1) delete all files in the /home/stevenr/import/ directory and /home/stevenr/EAC directory by issuing the following command in each:

verify directory by entering pwd, then enter:

```
rm *
```

ensure all files have been deleted by issuing the following command:

```
ls -l
```

Note: be sure this command is issued from /home/stevenr/import/

rm permanently deletes files and inadvertent deletions can cause permanent loss of data such as programming files that are critical to the import process.

- 2) Issue the following command from the command prompt in order to create the import files from home/stevenr:

```
perl eacimport.pl fonds.dmp boxFile.dmp ./EAC/eac
perl eadimport.pl fonds.dmp ./import/ead
```

- 3) Import EAC files from the CLI in /usr/share/nginx/atom:

```
sudo php5 symfony import:bulk --noindex="True"
/home/stevenr/EAC
```

- 4) Run the Index according to Step 4, then backup using

```
shell backup.sh
sudo mv <backupday>_Backup.sql atom_base-actors.sql
```

- 5) Once the EAC import is completed, import the EAD files

```
sudo php5 symfony import:bulk --noindex="True"
/home/stevenr/import
```

2 Post-import Procedures

AtoM contains deficiencies in the import functionality. It is not always possible to map fields in the input format directly to the fields displaying the information in the AtoM views. In some cases there is no input definition that corresponds to the fields displayed in AtoM. We have also observed input definitions that have an effect on more than one field. When this occurs the fields are sometimes unrelated causing unnecessary dependencies and side effects in the views. For example an authority record exported in the EAD format and subsequently re-imported has a different view than the original record. In order to overcome these issues post import procedures will be followed to correct the data at the source in the atom database.

The following issues were addressed:

1. Adjust location information

The location field cannot be populated through the ead import format. All location data is imported into the Name field and subsequently moved to the location field.

2. Fix anomalies in actor Roles

When importing the creator, accumulator roles all roles revert to creator even though the roles are correctly defined in the ead import file.

Corrections are made for these issues by executing the following command from the mysql prompt:

```
\. buildActors.sql
```

NOTE: refer to the Data Migration Project Developer Documentation for a more detailed explanation of the adjustments that were made.

3 Removing Untitled Creators/Accumulators

Determine the id for untitled authors by performing the following query:

```
select id, authorized_form_of_name from actor_i18n where
authorized_form_of_name like '%titled%';
```

in this example we will use id = 317

We wish to keep the untitled author wherever a date of creation exists therefore the following query deletes only untitled authors without a corresponding date of creation:

```
select event.id, event_i18n.date, event.actor_id from event left join
event_i18n ON event.id = event_i18n.id where (event_i18n.date is NULL
or event_i18n.date = " ") and event.actor_id = 317;

select count(*) from event;

select count(*) from event_i18n;
```

Counts must agree if referential integrity is maintained

This statement deletes the entries in events_i18n;

```
delete from event_i18n where id IN (select id from (select event.id
from event left join event_i18n ON event.id = event_i18n.id where
(event_i18n.date is NULL or event_i18n.date = " ") and event.actor_id
= 317) AS tmpstable);
```

To delete the corresponding entries in events enter the following command;

```
delete from event where id NOT IN (select id from event_i18n);
```

4 Search Indexing

In order to index items in AtoM, perform the following commands:

Clear cache from /usr/share/nginx/atom

```
sudo php5 ./symfony cc
```

Then index:

```
sudo php5 ./symfony search:populate
```

5 Publishing Archival Materials

In order to change the status of all materials from draft to published perform the following command from the mysql prompt:

```
update status set status_id = 160 where status_id = 159;
```


APPENDIX

Files Used

split.pl

orphanedfiles.pl

unlinkedFindingAids.pl

eacimport.pl

eadimport.pl

backup.sh

buildActors.sql