

Chapter 5

Importing data

EAD 2002

The Archivists' Toolkit allows you to import valid EAD version 2002 finding aids into description resource and component records.

Note: The AT does not guarantee complete round tripping of data. For more details, see the section on data mapping at the end of this chapter, and the EAD to AT data map in the appendices.

Constraints on EAD

To be imported, the EAD must meet the following conditions:

1. EADs must:
 - a. Be valid version 2002 documents. Version 1.0 EADs will have to be converted to version 2002 prior to import.
 - b. Contain a <unitid> within the <archdesc>
 - c. Conform to the EAD DTD (EADs using the EAD group DTD are not accommodated).
2. A valid EAD DTD (ead.dtd) for version 2002 must be present in the directory from which the EAD document is being imported
3. Not all valid EAD tagging can be accommodated by the Toolkit's data model, though such instances should be rare. Where inline tagging is encountered in a source EAD, it will be imported into the Toolkit as mixed content and will be visible as XML in its appropriate database field. Examples of inline content include formatting tags (such as <emph>, <lb/> and tables), access points outside of <controlaccess> tags, and external references (with the exception of dao types, which are mapped as digital instances). For example, a note encoded as follows:

```
. . .taken at the crossing of the Ebro in July
1938. In an article that appeared in the <title
render="italic">Volunteer for Liberty</title>
in September . . . Ebro Offensive:</p>
```

```
<blockquote><p>
```

```
In the period of reorganization ...
```

```
</p></blockquote>
```

Will be imported with some tagging retained as mixed content:

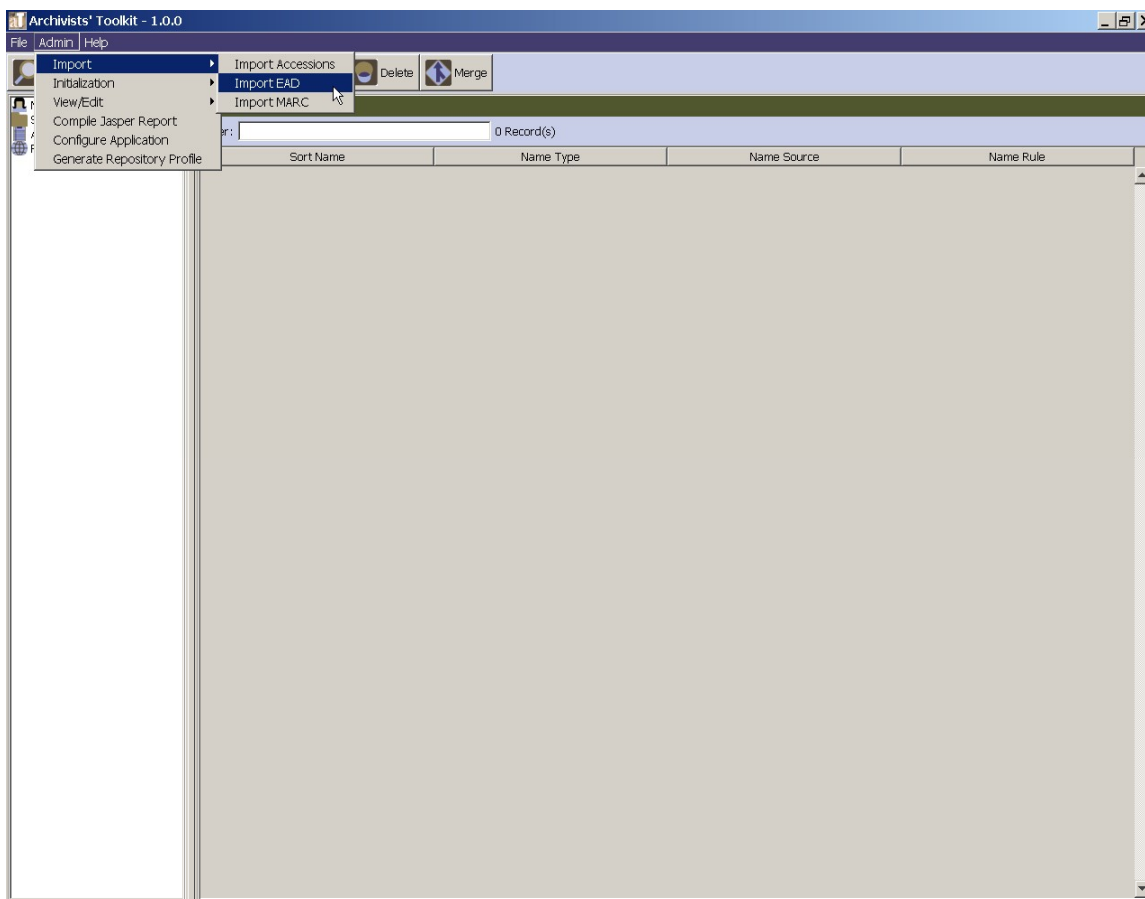


Note: As you will see in the example above, paragraphs are displayed in the AT with two line returns. Therefore, the `<p>` tags do not appear in the note.

4. Most internal referencing between id and target tag attributes will be retained with some provisions: the Toolkit's "persistent identifier" will replace the id target pairing and the Toolkit's "targets" will be limited to names, notes, and components.
5. The Toolkit does not currently importing a digital object title or objectType for `<dao>` elements in EADs even though these fields are required in the Toolkit. If you are importing EADs containing `<dao>` elements, you may wish to add title and objectType information to these records in the Toolkit so that the resulting digital object records are valid. (It is possible, however, to export these as `<dao>` elements in EAD even if the digital object is invalid in the Toolkit.)

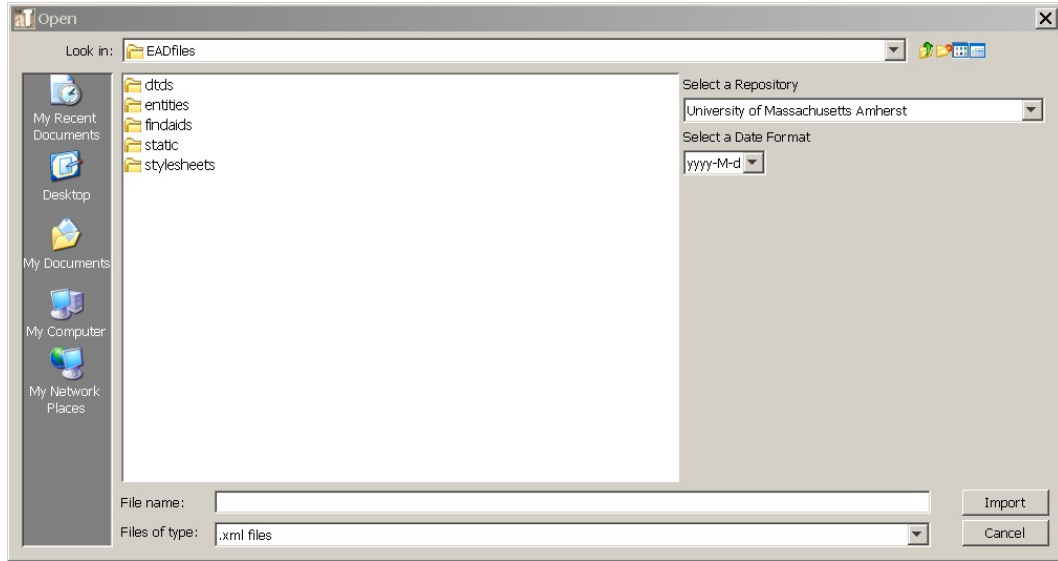
Importing EAD

1. From the **Admin** menu, select **Import** and **Import EAD**.



2. Choose the file to be imported.
3. Select the repository that holds the materials described in the finding aid

4. Indicate normalized date format.



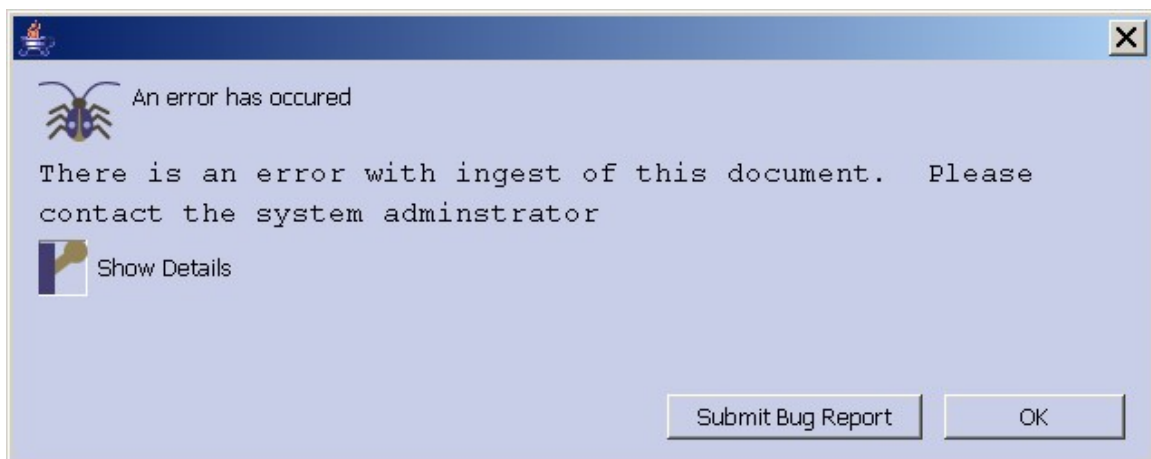
5. Click import.

The Toolkit will display a graphic indicating the progress of the import:

The screenshot shows the Archivists' Toolkit interface. The main window displays a list of records with columns for Sort Name, Name Type, Name Source, and Name Rule. A loading spinner is overlaid on the list, and the text "Loading finding aid into memory" is visible at the bottom of the spinner.

Sort Name	Name Type	Name Source	Name Rule
[Persname] Creator, Author or:	Person	ingest	
Capit. c. Vladimir, 1891-1939.	Person	ingest	
Aalto, William.	Person	ingest	
Actors' Equity Association.	Corporate Body	ingest	
Adam, Helen., 1909-	Person	ingest	
Adler Family	Family		Anglo-American Cataloguing Rules, 2nd ed.
Adler, Irving	Person	ingest	
AFSCME sdf adsfads f.	Corporate Body	ingest	
Albany Argus..	Corporate Body	ingest	
Allen, Judson Boyce, 1932-	Person	ingest	
Allen, William, 1784-1868	Person	ingest	
Alleyn's College of God's Gift (London, Engl...	Corporate Body	ingest	
American Architectural Foundation	Corporate Body	ingest	
American Association of University Profess...	Corporate Body	ingest	
American Land Company.	Corporate Body	ingest	
American Medical Research Expedition to ...	Corporate Body	ingest	
American Philosophical Society (Philadelph...	Corporate Body	ingest	
American Red Cross.	Corporate Body	ingest	
Anderson, Henry, d. 1846.	Person	ingest	
Anderson, Peter, 1804-1868.	Person	ingest	
Anderson, Walter Alexander, 1909-	Person	ingest	
Andron Club of New York City..	Corporate Body	ingest	
Andrews, Bruce., 1948-	Person	ingest	
Anti-Fascisti di Staten Island--Pictorial wor...	Corporate Body	ingest	
Aptheker, Herbert, 1915-	Person	ingest	
Archdiocese of Santa Fe (Albuquerque, N...	Corporate Body	ingest	
Arion-New York.	Corporate Body	ingest	
Arquidiocesis del Cusco (Cusco, Peru)	Corporate Body	ingest	
Art Institute of Chicago (Chicago, Illinois)	Corporate Body	ingest	
Ashbery, John,	Person	ingest	
Asian Art Museum Foundation of San Fran...	Corporate Body	ingest	
Asiatic Society of Bangladesh (Bangladesh)	Corporate Body	ingest	
Asociación Amigos de Salas Nacionales y ...	Corporate Body	ingest	
Auckland Museum Trust Board (Auckland, ...	Corporate Body	ingest	
Ayers, Stephen B.	Person	ingest	
Azarias, Brother., 1847-1893.	Person	ingest	
Bailey, Bill.	Person	ingest	
Baliti Heritage Trust (Karimabad, Pakistan)	Corporate Body	ingest	
Barbados Museum and Historical Society, ...	Corporate Body	ingest	
Barone, Dennis,	Person	ingest	
Barry, N.C.	Person	ingest	
Bay Area Video Coalition, Inc. (San Francisc...	Corporate Body	ingest	
Beland, Pura, 1909-1982.	Person	ingest	
Benson, C. David	Person	ingest	
Bernard, Theodore C., 1903-	Person	ingest	
Bernstein, Charles., 1950-	Person	ingest	
Berol, Alfred C. (collector).	Person	ingest	
Bertholf, Robert J.,	Person	ingest	
Bessie, Alvah Cecil, 1904-	Person	ingest	
Bessie, Alvah Cecil, 1904--Death and buri...	Person	ingest	
Bessie. Nan.	Person	ingest	

If an error, such as an invalid EAD is encountered, the Toolkit will display an error message indicating that the file could not be imported:

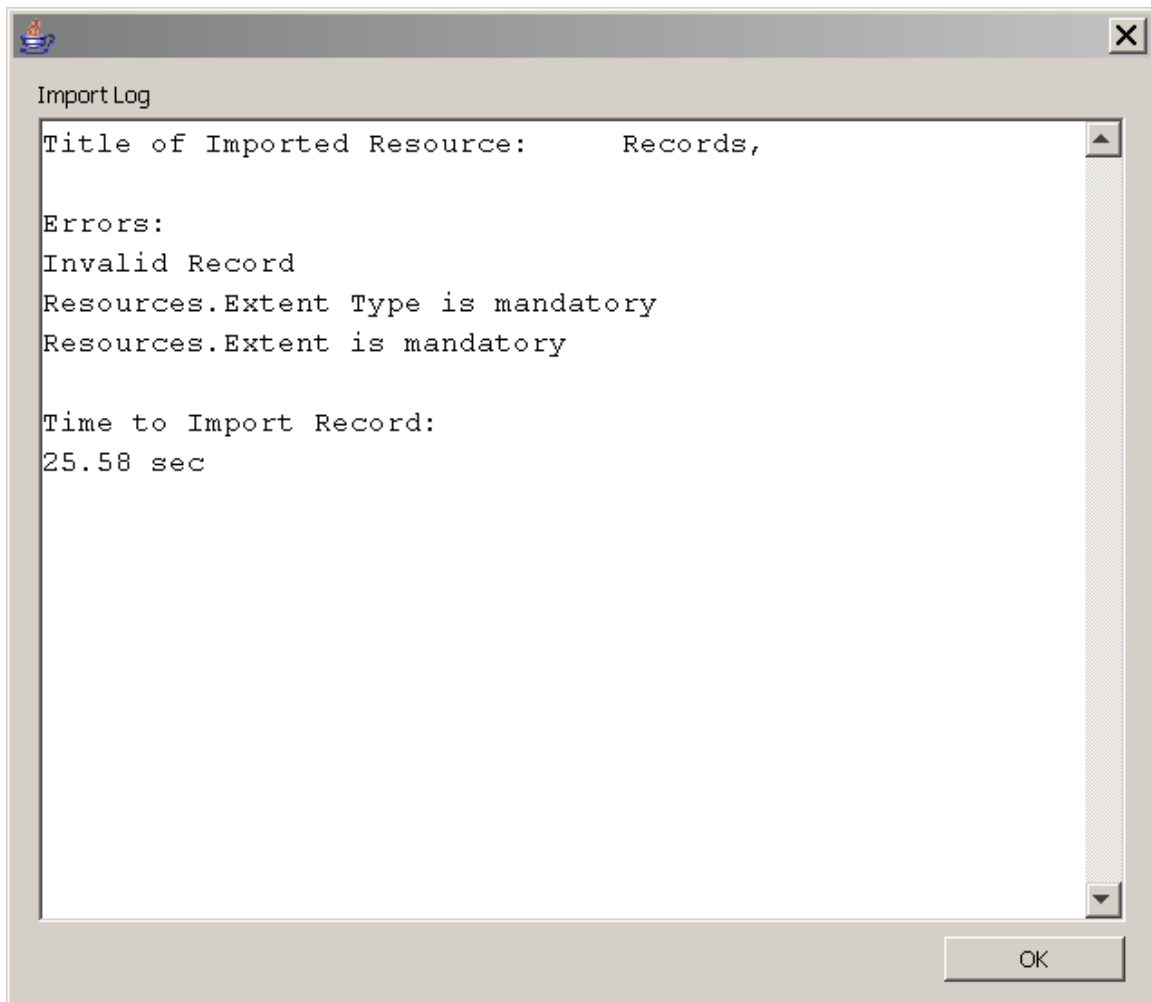


There are several factors that could result in a failed import of an EAD instance:

- The EAD instance is not a version 2002 EAD

- The EAD instance is not well formed
- The EAD to be imported already exists in the AT as a multi-level resource, that is there is already a resource record in the AT with the same resource identifier as the EAD to be imported and that resource record has linked resource component records. If the resource record does not have linked resource component records, then the EAD to be imported will be imported and will overwrite the resource record that is already there. (This principle is largely the same for ingesting parent level MARC records, as well.)

If import of the record is successful, an import log will appear, listing any errors encountered in the import process:



These errors typically consist of absent data elements that are necessary for a valid AT record. You will be prompted to correct these errors once you open the resulting record in the AT. They must be corrected in order for you to save the record again.

Mapping of EAD elements to Archivists' Toolkit fields

Note: A complete EAD to AT mapping is available in the Appendices. It is organized into four sections, each on an individual tab in the Excel spreadsheet. Though some aspects of the mapping are spelled out below, those interested in the complete details of how EAD elements are imported into the Toolkit should refer to the mapping document.

The entire EAD data structure does not map directly to the AT's data structure, so in order to make sure that no data is lost, certain elements will be imported as mixed content. That is, some data will be imported into a field with all of its internal tagging intact. Tables and linking elements are examples of EAD elements that are imported as mixed content. For example, a biographical note (<bioghist>) that contains a <table> will be imported into the Toolkit's biographical/administrative history note field, with all content and tagging.

Linking elements will also be imported as mixed content. For internal references imported into the Toolkit (<ptr>, <ptrloc>, <ref>, <refloc>) the target and id pairs will be realigned so that they contain Toolkit persistent ID values. For descriptions created within the AT, only <extref> and <ref> elements are used by the **Wrap in tag** editor, though you are free to hand code any other linking elements you wish to use.

A final note on linking elements relates to importing indexes. Within an <index>, the Toolkit only supports one link per index entry. Therefore, <ptrgrp> elements are not imported.

MARC XML

The Archivists' Toolkit allows you to import valid MARCXML records into description resource records; however, the AT is not designed to manage MARC records. The MARC import function is provided simply to allow repositories to bring all their resources, be they described in legacy MARC records, EAD records, or created from scratch in the AT under management within a single tool.

<p>Caution: The AT MARC import function does not support round tripping of MARC records (i.e., the MARCXML record exported from the AT will not be the same as the MARCXML record that was imported). The import of MARC records into the AT involves loss of data granularity, and in some cases, loss of data. For instance, some subfield demarcations are not supported in the AT; so, although the data is imported, it is concatenated within the same AT field. And, while the AT is designed to support the MARC data elements typically used for describing archival materials, there may be cases where an imported MARC record contains a field not supported by the AT.</p>

Despite these constraints, repositories may wish to import MARCXML in the following scenarios:

- When legacy data for archival resources is in the form of MARC records and not in the form of EAD finding aids and the repository wants to control the resources described in the MARC records in the AT.
- When the MARC records for archival materials are considered to contain the only, or the more authoritative, name and subject headings than what are present in a repository's EAD finding aids.

Import has been intentionally designed as a “one record at a time” process based on the aforementioned recommendations for MARCXML import scenarios. Though MARCXML records usually are exported from an ILS as a “batch,” importing batches of resources may cause AT records to be unintentionally overwritten or duplicated.

Constraints on MARCXML

To be imported, the MARCXML record must meet the following conditions:

1. Valid MARCXML record with type coded to “bibliographic.”
2. The source file must contain only a single MARCXML record.
3. The MARCXML must be for a resource under archival descriptive control (e.g. appm, DACS, etc.).
4. Top level records. MARC records that contain a 773 field, indicating the record is a “child” record related to a “parent” record, will not be imported.
5. The MARC record must contain a 210, 222, 240, or 245 title field or the record will not be imported.
6. If the tagging follows the <marc:[tag]> convention, the namespace declaration of the imported document must be:

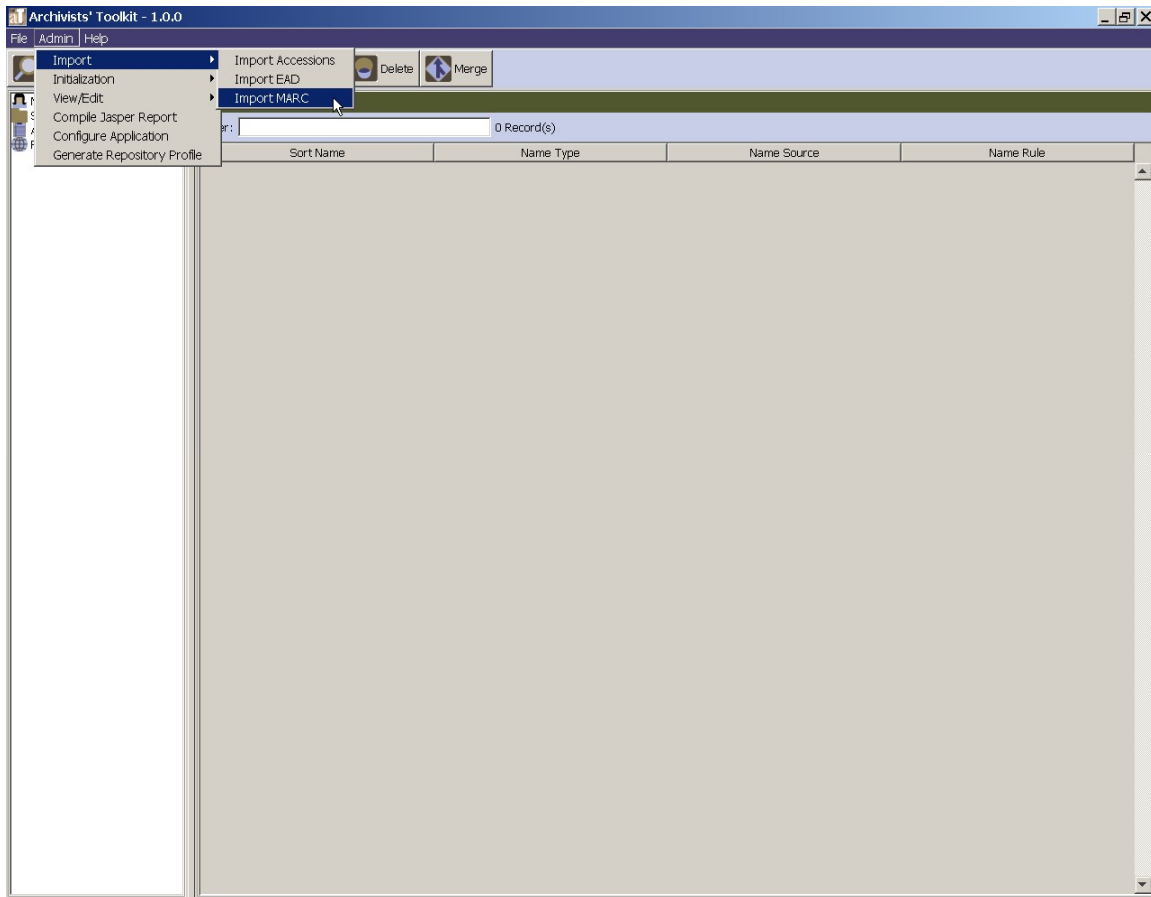
```
xmlns:marc="http://www.loc.gov/MARC21/slim"  
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation=http://www.loc.gov/MARC21/slim  
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd
```

Otherwise the namespace declaration must be:

```
xmlns="http://www.loc.gov/MARC21/slim"  
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation="http://www.loc.gov/MARC21/slim  
http://www.loc.gov/standards/marcxml/schema/MARC21.xsd"
```

Importing MARCXML

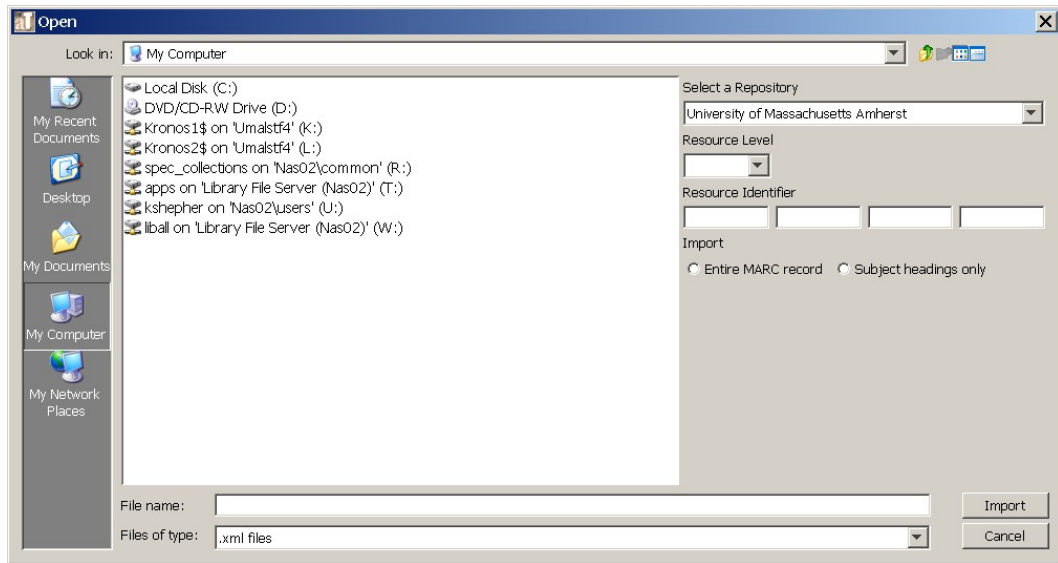
1. From the **Admin** menu, select **Import**, then **Import MARC**.



2. Choose the file containing the MARCXML record to import.
3. Select the repository that holds the materials described in the MARC record.
4. Indicate resource identifier.

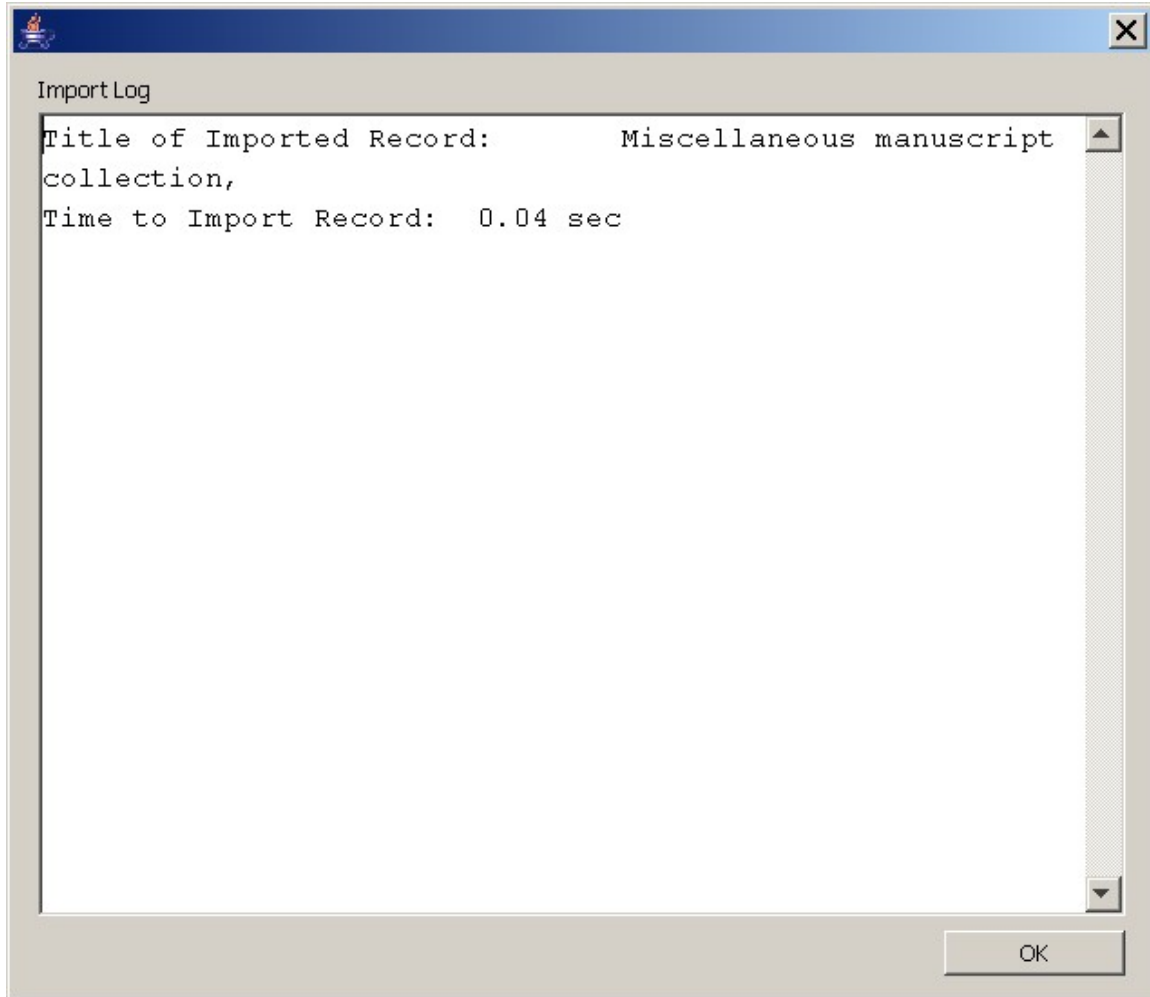
Note: If the resource identifier matches the resource identifier of an existing record, the Toolkit will ascertain whether or not the existing record contains linked component records. If it does not contain linked component records, the resource record will be overwritten. If it does contain linked component records, the Toolkit will respond that the MARC XML record cannot be imported.

5. Indicate whether you wish to import the entire MARCXML record or ONLY the 1xx, 6xx, and 7xx fields for name and subject headings.



6. Click import.

If an error, such as an invalid MARCXML is encountered, the Toolkit will display an error message indicating that the file could not be imported. If import of the record is successful, an import log will appear, listing any validation errors present in the imported record:



These errors typically consist of absent data elements that are necessary for a valid AT record. The first time you open the resulting AT record the AT will prompt you to correct these errors, and will not allow you to save the record again until the corrections have been made.

Mapping of MARCXML elements to Archivists' Toolkit fields

A MARCXML to AT mapping is available in the appendices.

Tab-delimited accessions data

Importing accessions data

The Archivists Toolkit allows you to import data in tab-delimited format from a table or flat file. Tab-delimited data may be imported into the accessions, names, or subjects functional areas. Before importing data into the Toolkit, the data must be mapped to the fields in the Toolkit. Some mapping tips are listed below.

Constraints on legacy accessions data

There are constraints on the format of import data that may require cleanup of your legacy data prior to import. Importing data that does not meet with these constraints prevent the record from being or cause errors in the resulting accession records in the Toolkit.

1. Data must be in a simple flat table; relational or linked tables will not be accommodated. The table should be tab delimited and unformatted text.
2. The Toolkit allows only one name to be imported with the accession record, and only one type of subject term (see below).
3. The fields of the import table must be identical to the fields listed in the mapping table. The only exception is the accessionNumber, which may have a one-to-many, source to target mapping (see #6 below).
4. If starting from a data table, such as one created by MS Access or FileMaker Pro, hard returns and other formatting within the cells of the table should be removed. Line breaks or hard returns in the data may be pipe delimited (|) and those pipe characters will be converted to line breaks in the Toolkit if the target field has a free-text data type.
5. Boolean fields (cataloged, restrictionsApply, rights) may only contain values for true or false.
6. ExtentNumber must be a number. Fractions must be expressed as decimal values and may not contain non-numeric characters.
7. When opening an imported accession record in the Toolkit, you'll be prompted to fix accession records that you wish to save that are missing required data.
8. Accession Number. The Toolkit offers two ways to import accession number data:
 - a. If the import field is labeled accessionNumber, the number will be parsed so that an accession number, 2005-54.1 will be targeted to accessionNumber1=2005, accessionNumber2=54, accessionNumber3=1.
 - b. You may bypass this parsing by labeling the import field containing 2005-54.1 as accessionNumber1.

Note: Duplicate accession numbers are not allowed and will cause an error

9. If the import field is labeled resourceIdentifier, the ID will be mapped so that a resource identifier, MC 543 will be targeted to resourceIdentifier1=MC 543 and stored as part of the resource record, if the option to create resource data is selected at the beginning of the import process (see below).
10. Dates

- a. **dateBegin** and **dateEnd** fields will accept only integer dates. A.D. dates should be expressed with four digits, e.g., 1972, 1514, 2006. B.C. dates should be expressed with a negative sign first, e.g., -999, -10000, -3.
- b. Complex date strings, such as ca. 1976-spring 1987, can be imported into the **dateExpression** field.

Note: the dateExpression field has a parser that will parse 1963-1987 into dateBegin=1963 and dateEnd=1987. Import of dateBegin without dateEnd will force you to enter the dateEnd when viewing the record in the Toolkit.

- c. Upon import, the Toolkit requires that you declare the formal date structure for transaction date fields (accessionDate, acknowledgementDate, agreementSent and agreementReceived). These transactional dates may be in the following formats: yyyy-mm-dd (example: 1963-09-27), yyyy/mm/dd (example: 1963/09/27), mm-dd-yyyy (example: 09-27-1963) or mm/dd/yyyy (example: 9/27/1963). Any data in this type of date field that is not properly formatted will cause an error.

11. Subject terms

- a. Importing one or more subjectTerms, requires the presence of a single subjectTermType [(Function (657), Genre / Form (655), Geographic Name (651), Occupation (656), Topical Term (650), Uniform Title (630)].
- b. Multiple subjectTerms may be imported by delimiting terms with a pipe character, for example: Boating. | United States History Civil War, 1861-1865. When multiple subjects are imported, they will all be given the same SubjectTermType value, as only one subjectTermType may be used for a row of data. Therefore, only one type of subject may be imported for an accession record. To overcome this limitation, you may choose to import accession data without subject headings and either import the subject headings separately or create them directly in the AT. Subject headings imported separately or created directly in the AT can be manually linked to appropriate accession records. Subject headings, especially multiple subject headings, imported with accession data will likely require remediation after import is complete. Repositories will need to determine which strategy is most efficient and cost-effective for them.
- c. On import, subject terms are assigned the source value import. You may then more easily identify subject terms that require editing or remediation.

12. Names. Only one name per accession may be imported, therefore you may elect to import only sources, only creators, or only the first subject name.

- a. Every name must have a Function (source, creator, or subject) and a corresponding primary name.

- b. If names exist in a single field in the legacy data, they may be imported into one of the Primary name fields, and parsed into individual fields by hand in the Toolkit. To load names into the correct type of primary name field, you must first separate each name into its appropriate target field, as shown below:

nameFunction	Name_corporatePrimary	Name_personalPrimary	Name_familyName
creator		Smith, John, 1904-1990	
subject			Smith Family
source	Acme Industrial Products		

Examples and Mapping

The following shows a sample import file in tab delimited format. Every space depicted is a single tab:

```
resourceIdentifier      accessionDate  accessionNumber extentNumber
      extentType      title
MSS 029 1957-01-01    1957.029      45      Linear feet    Fales
Family Papers
MSS 002 1964-01-01    1964.002      126     Linear feet    The
Elizabeth Robins Papers
```

The following table shows the allowed field names and the target fields for accession data to be imported into the Toolkit. Some fields have controlled vocabularies, and those are listed in the Rules column.

Import field/column label	Where stored /AT target	Data Type	Rule
accessionDate	accessionDate	Date	
accessionDispositionNote	accessionDispositionNote	Text	
Import field/column label	Where stored /AT target	Data Type	Rule
accessionNumber		Text	parsed by delimiters into accessionNumber1, accessionNumber2, accessionNumber3, and accessionNumber4
accessionNumber1	accessionNumber1	Text	
accessionNumber2	accessionNumber2	Text	
accessionNumber3	accessionNumber3	Text	
accessionNumber4	accessionNumber4	Text	
accessRestrictions	accessRestrictions	Boolean	
acknowledgementDate	acknowledgementDate	Date	
acquisitionType	acquisitionType	Text	Deposit, Gift, Purchase, Transfer
agreementReceived	agreementReceived	Date	
agreementSent	agreementSent	Date	
bulkDateBegin	bulkDateBegin	Integer	

bulkDateEnd	bulkDateEnd	Integer	
cataloged	Cataloged	Boolean	
catalogedNote	catalogedNote	Text	
conditionNote	conditionNote	Text	
containerSummary	containerSummary	Text	
dateBegin	dateBegin	Integer	
dateEnd	dateEnd	Integer	
dateExpression	dateExpression	Text	
dateProcessed	dateProcessed	Date	
description	Description	Text	
extentNumber	extentNumber	Integer	
extentType	extentType	Text	Linear Feet, Cubic Feet
inventory	Inventory	Text	
processingPlan	processingPlan	Text	
resourceType	resourceType	Text	Collection, Papers, Records
restrictionsApply	restrictionsApply	Boolean	
retentionRule	retentionRule	Text	
rightsTransferred	rightsTransferred	Boolean	
title	title	Text	
		Text	

resourceIdentifier

For import of names from an accession table, only one name can be imported at this point. The name must contain: nameRole. If no nameRole is specified, then default nameRole=subject.

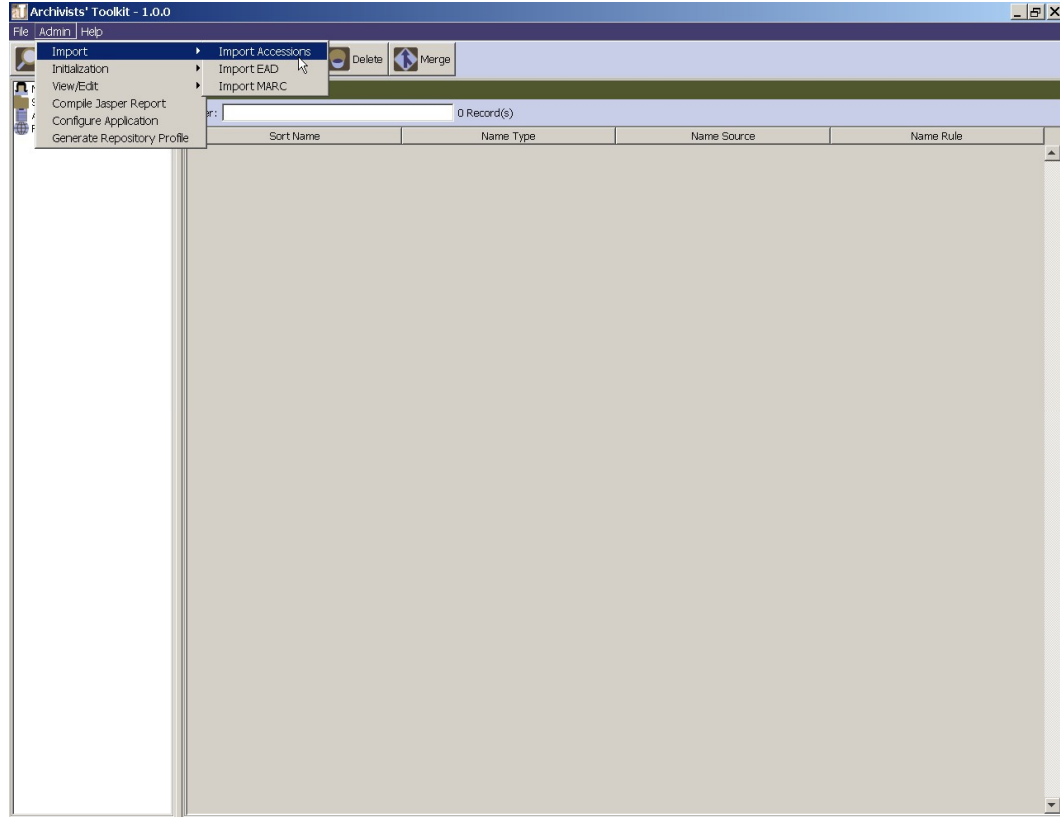
Name_contactAddress1	contactAddress1	Text	
Name_contactAddress2	contactAddress2	Text	
Name_contactCity	contactCity	Text	
Name_contactCountry	contactCountry	Text	
Name_contactEmail	contactEmail	Text	
Name_contactFax	contactFax	Text	
Name_contactMailCode	contactMailCode	Text	
Name_contactName	contactName	Text	
Name_contactPhone	contactPhone	Text	
Name_contactRegion	contactRegion	Text	
Name_corporateNumber	number	Text	
Name_corporatePrimaryName	primaryName sortName	Text	
Name_corporateQualifier	qualifier	Text	
Import field/column label	Where stored /AT target	Data Type	Rule
Name_corporateSubordinate1	corporateSubordinate1	Text	
Name_corporateSubordinate2	corporateSubordinate2	Text	

Archivists' Toolkit User Manual

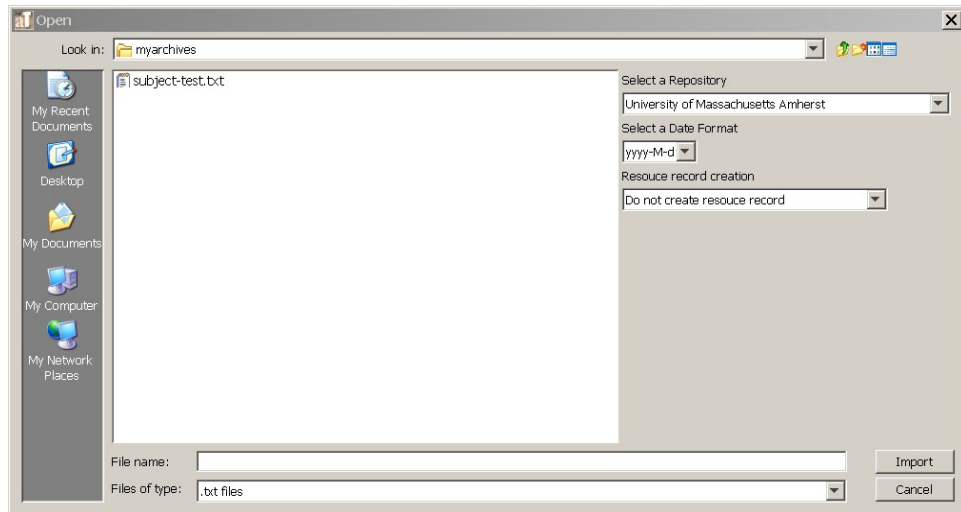
Name_descriptionNote	descriptionNote	Text	
Name_descriptionType	descriptionType	Text	
Name_familyName	PrimaryName	Text	
Name_familyNamePrefix	namePrefix	Text	
Name_familyQualifier	qualifier	Text	
Name_personalDates	dates	Text	
Name_personalFullerForm	fullerForm	Text	
Name_personalNumber	number	Text	
Name_personalPrefix	prefix	Text	
Name_personalPrimaryName	primaryName sortName	Text	
Name_personalQualifier	qualifier	Text	
Name_personalRestOfName	restOfName	Text	
Name_personalSuffix	suffix	Text	
Name_Source	source	Text	
nameFunction	nameFunction	Text	Creator, Souce, Subject
		Text	
		Text	Multiple subjects will be accommodated, see conditions below. Each term must have a subjectTermType.
		Text	If more than one subjectTerm is listed in pipe delimited format, the subjectTermType will be the same for each subject listed. Types are: Function (657), Genre / Form (655), Geographic Name (651), Occupation (656), Topical Term (650) ,Uniform Title (630)
subjectTermType	subjectTermType	Text	
subjectTerm	subjectTerm	Text	pipe delim for multiple
subjectSource	subjectSource	Text	default "import "

Importing accessions data

1. From the Admin menu, select Import and Import Accessions.



2. Make the following selections:
 - a. Choose the file to be imported.
 - b. Linked resource records:
 - Create no resource records linked to the imported accession records
 - Create linked resource records containing only the resource identifier imported from the accession table.
 - Create more complete linked resource records, utilizing imported accession data in the following fields: extent, container summary, date, title, names, subjects, rights transferred, access restrictions.
 - c. Select the repository to which the accession data applies.
 - d. Indicate the format of the dates (see constraints) contained in the file to be imported.



3. Click import to begin.

A progress bar will track the import process. If errors are encountered, an error log will appear at the end of the process.

Note: Several error conditions can cause the import to fail, in whole or in part:

- An improperly formatted import document will cause the entire process to fail. No records will be imported
 - An invalid record will not be imported. To be valid an accession record must have an accession number and an accession date.
 - Incorrectly formatted data, e.g., a non-accepted date format, or data mismatch, e.g., text data where integer data is required, will cause import of a record to fail.
-

If the import document is formatted correctly, the process will proceed record by record. All valid and error free records will be imported. All invalid and / or errant records will not be imported and a list of them will be created.

If no errors are encountered, you will see a window like the one shown directly below.

