

Appendix: Preparing XML Accession Files

Importing XML accessions data

The Toolkit will not import accessions information straight from a database, such as Access or Filemaker Pro. One method of importing accessions data is through the use of an XML file based on the XML schema included with the Toolkit. The XML format supports the import of multiple names and subjects types. The XML schema needed to create XML formatted data is named `accessionsImport.xsd` and can be found in the “conf” folder in the directory where the Toolkit is stored. Before importing data into the Toolkit, the source accessions data must be mapped to the fields supported by the AT accession import schema. Some mapping tips are listed below.

There are many ways to configure accessions data to the Toolkit’s XML schema. All require some understanding of XML. One method is to export the accessions database information into XML and write an XSLT stylesheet to map the database fields to the Toolkit’s schema. Another method is to duplicate the database (for backup purposes), replace the original database fields with the mapped AT fields (listed below), and export the database as an XML file. Other options are possible depending on what type of technical knowledge is accessible to your repository.

Element Mapping

The following table shows the allowed elements for accession data to be imported into the Toolkit. Some fields have controlled vocabularies, and those are listed in the **Rule** column. The **Field Length** is the maximum number of characters allowable for each field. The XML will not validate against the Toolkit schema if field lengths are too long. **AccessionNumber** and **accessionDate** must be included in the XML for each accession for the file to validate. Fields need to be listed in the same order as they are presented in the `accessionsImport.xsd` file. Examples of XML import files are included at the end of this chapter.

Accession Elements			
<u>AT Element</u>	<u>Field Length</u>	<u>Data Type</u>	<u>Rule</u>
accessionNumber		text	parsed by delimiters into: accessionNumber1 accessionNumber2 accessionNumber3 accessionNumber4
accessionDate		date	
accessionDispositionNote	no limit	text	
accessRestrictions		boolean	
accessRestrictionsNote	no limit	text	
useRestrictions		boolean	
useRestrictionsNote		text	
generalAccessionNote	no limit	text	
acknowledgementDate		date	

acknowledgementSent		boolean	
agreementReceivedDate		date	
agreementSentDate		date	
agreementSent		boolean	
agreementReceived		boolean	
accessionProcessed		boolean	
accessionProcessedDate		date	
acquisitionType	255	string	Deposit, Gift, Purchase, Transfer
dateExpression	255	string	
bulkDateBegin		integer	
bulkDateEnd		integer	
dateBegin		integer	
dateEnd		integer	
description	no limi	text	
cataloged		boolean	
catalogedDate		date	
catalogedNote	no limit	string	
conditionNote	no limit	string	
extentNumber		real number	
extentType	255	string	Linear Feet, Cubic Feet
containerSummary	no limit	string	
inventory	no limit	text	
processingPlan	no limit	text	
processingPriority	255	string	
processors	255	string	
processingStatus	255	string	
processingStartedDate		date	
resourceType	255	string	
retentionRule	no limit	text	
title	255	string	
restrictionsApply		boolean	
rightsTransferred		boolean	
rightsTransferredDate		date	
rightsTransferredNote	no limit	text	
userDefinedDate1		date	
userDefinedDate2		date	
userDefinedBoolean1		boolean	
userDefinedBoolean2		boolean	
userDefinedInteger1		integer	
userDefinedInteger2		integer	
userDefinedReal1		real number	
userDefinedReal2		real number	
userDefinedString1	255	string	
userDefinedString2	255	string	
userDefinedString3	255	string	

userDefinedText1	no limit	text	
userDefinedText2	no limit	text	
userDefinedText3	no limit	text	
userDefinedText4	no limit	text	
resourceIdentifier	20	string	
<u>Where Stored/AT Target</u>	<u>Field Length</u>	<u>Data Type</u>	<u>Rule</u>
subjectTerm	150	string	
subjectTermType	50	string	Types are: Function (657), Genre / Form (655), Geographic Name (651), Occupation (656), Topical Term (650), Uniform Title (630)
subjectSource	100	string	default "ingest "
<u>Where Stored/AT Target</u>	<u>Field Length</u>	<u>Data Type</u>	<u>Rule</u>
nameLinkFunction	50	string	Creator, Source, Subject
nameType	255	string	Person, Corporate Body, Family
sortName	255	string	
number	255	string	
qualifier	255	string	
nameSource	50	string	
nameRule	50	string	
descriptionType	255	string	Administrative History, Biography
descriptionNote	no limit	text	
citation	no limit	text	
salutation	255	string	
contactAddress1	255	string	
contactAddress2	255	string	
contactCity	255	string	
contactRegion	255	string	
contactCountry	255	string	
contactMailCode	255	string	
contactPhone	255	string	
contactFax	255	string	
contactEmail	255	string	
contactName	255	string	
familyName	35	string	
familyNamePrefix	15	string	
corporatePrimaryName	70	string	
corporateSubordinate1	45	string	
corporateSubordinate2	45	string	
personalPrimaryName	30	string	
personalRestOfName	25	string	
personalPrefix	25	string	
personalSuffix	25	string	
personalDates	15	string	
personalFullerForm	25	string	
personalTitle	25	string	

personalDirectOrder		boolean	
---------------------	--	---------	--

Constraints on legacy accessions data

There are constraints on the importing of XML format of accessions data in XML format that may require cleanup of your legacy data prior to import. Not meeting these constraints prevents the record from being ingested into the Toolkit.

The following steps are required for correct accession data import:

1. The XML accession file must validate against the schema:
 - a. Dates have to be in valid XML date format, i.e. yyyy-mm-dd.
 - b. Boolean fields (**cataloged**, **restrictionsApply**, **rights**) may only contain the values true, false or 1, 0.
 - c. Integer fields (**dateBegin**, **dateEnd**) can only contain whole numbers (for example, 2004, or -50).
 - d. Real number fields (**extentNumber**) can contain integers and/or decimals.
2. If you have a single field in the database that needs to get mapped into more than one field in the Toolkit, you will need to parse the data into two separate fields. For example, in many cases accessions data will combine extent into one field, such as “11 linear feet”. The Toolkit provides two separate fields for extent: **extentNumber** and **extentType**. Extent number contains the numerical value of the extent (as a real number) and extent type contains the extent measurement type of the accession, such as linear feet.
3. 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 parsed out as **accessionNumber1**=2005, **accessionNumber2**=54, **accessionNumber3**=1. Periods, forward slashes, and dashes are the three characters that will delimit the accession number fields.
 - b. You may bypass this parsing by labeling the import field containing 2005-54.1 as **accessionNumber1**, causing the accession number to be stored in one field. Be aware that the each of the four accession number fields are limited to ten characters.

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

4. If the import field is labeled **resourceIdentifier**, the ID will be mapped so that a resource identifier, MC 543 will be imported as **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).
5. Dates

- a. **dateBegin** and **dateEnd** fields will accept only integer dates. A.D. dates after 1000 A.D. should be expressed with four or less digits, e.g., 814, 1514, 2006. B.C. dates should be expressed with a negative sign first, e.g., -999, -10000, -3.
- b. Natural language 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**. If a **dateBegin** is provided without a corresponding **dateEnd**, or if the **dateEnd** field is less than the **dateBegin** field, the record will not be imported.

6. Subjects. Multiple subjects are allowed. Each **subjectTerm** must have a corresponding **subjectTermType** and **subjectSource**. If no **subjectSource** is included, the system assigns the value “ingest” to that element. The following is an example of how subjects are represented according to the XML schema:

```
<subjectLink>
  <subjectTerm>Railroads -- Mexico</subjectTerm>
  <subjectType>Topical Term (650)</subjectType>
  <subjectSource>Library of Congress Subject Headings
</subjectSource>
</subjectLink>
```

7. Names

- a. Every name must have a **nameFunction** (source, creator, or subject) and a corresponding primary name. The function describes the relationship of the name to the accession. The name can additionally have a Role and a Form.
- b. Every name must also have a **nameType**. The **nameTypes** are: Person, Corporate Body or Family.
- c. 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 in the name example below:

```
<nameLink>
  <nameLinkFunction>Source</nameLinkFunction>
  <name>
    <nameType>Person</nameType>
    <nameSource>lcnaf</nameSource>
    <nameRule>DACS</nameRule>
    <contactAddress1>111 Broadway</contactAddress1>
    <contactCity>New York</contactCity>
    <contactRegion>New York</contactRegion>
    <contactCountry>US</contactCountry>
    <contactMailCode>11211</contactMailCode>
    <contactEmail>don.clark@gmail.com</contactEmail>
    <personalPrimaryName>Clark</personalPrimaryName>
```

```
        <personalRestOfName>Don R.</personalRestOfName>
    </name>
</nameLink>
```

XML Accession File Examples

Example One: Simple Accession Files

```
<?xml version="1.0" encoding="UTF-8"?>
<accessionRecords xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="file:/Users/sibylroud/Desktop/accessionsImport.xsd">
  <record>
    <accessionNumber>
      <part1>MSS</part1>
      <part2>2007</part2>
      <part3>034</part3>
    </accessionNumber>
    <accessionDate>2007-08-14</accessionDate>
    <extentNumber>20</extentNumber>
    <extentType>Linear feet</extentType>
    <title>Henry Jones papers</title>
  </record>
</accessionRecords>
```

Example Two: Accession Files with Subjects and Names

```
<?xml version="1.0" encoding="UTF-8"?>
<accessionRecords xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="file:/Users/sibylroud/Desktop/accessionsImport.xsd">
  <record>
    <accessionNumber>
      <part1>MC</part1>
      <part2>210</part2>
    </accessionNumber>
    <accessionDate>2007-12-13</accessionDate>
    <extentNumber>.5</extentNumber>
    <extentType>Linear feet</extentType>
    <containerSummary>programs; participant portfolios; participant bios</containerSummary>
    <title>Records of the Global Colloquium of University Presidents on Climate Change</title>
    <restrictionsApply>>false</restrictionsApply>
    <subjectLink>
      <subjectTerm>Climate change</subjectTerm>
      <subjectTermType>Topical Term (650)</subjectTermType>
      <subjectSource>Library of Congress Subject Headings</subjectSource>
    </subjectLink>
    <subjectLink>
      <subjectTerm>University presidents</subjectTerm>
    </subjectLink>
  </record>
</accessionRecords>
```

```

    <subjectTermType>Occupation (656)</subjectTermType>
    <subjectSource>local</subjectSource>
</subjectLink>
<nameLink>
  <nameLinkFunction>Source</nameLinkFunction>
  <name>
    <nameType>Person</nameType>
    <nameSource>local</nameSource>
    <nameRule>DACS</nameRule>
    <contactAddress1>Office of University Relations and Public Affairs</contactAddress1>
    <contactAddress2>111 Broadway</contactAddress2>
    <contactCity>New York</contactCity>
    <contactRegion>New York</contactRegion>
    <contactCountry>US</contactCountry>
    <contactMailCode>11211</contactMailCode>
    <contactEmail>Beyeni@nyu.edu</contactEmail>
    <personalPrimaryName>Beyeni</personalPrimaryName>
    <personalRestOfName>Amelia</personalRestOfName>
  </name>
</nameLink>
</record>
</accessionRecords>

```

Example Three: Accession Files with Subject, Names, and User-Defined Fields

```

<?xml version="1.0" encoding="UTF-8"?>
<accessionRecords xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="file:/Users/sibylroud/Desktop/accessionsImport.xsd">
  <record>
    <accessionNumber>
      <part1>RG</part1>
      <part2>20</part2>
      <part3>9</part3>
    </accessionNumber>
    <accessionDate>2007-08-09</accessionDate>
    <accessRestrictions>true</accessRestrictions>
    <accessRestrictionsNote>This collection may contain materials pertaining to student records or
student work. Viewing of such materials may be prohibited under FERPA guidelines. Check with the
University Archivist for specific restriction guidelines.</accessRestrictionsNote>
    <generalAccessionNote>Little to no discernible order</generalAccessionNote>
    <accessionProcessed>>false</accessionProcessed>
    <dateExpression>ca. 1985-2000</dateExpression>
    <extentNumber>13</extentNumber>
    <extentType>Linear feet</extentType>
    <inventory>Student work; published articles; newsletters; planning materials; memoranda;
emails; exhibit materials</inventory>
    <processingStatus>new</processingStatus>
    <title>Records of the Public History Program</title>
    <userDefinedBoolean1>true</userDefinedBoolean1>

```

```
<userDefinedInteger1>6</userDefinedInteger1>
<userDefinedText1>Materials picked up from the Public History Office in King Juan Carlos I in
July 2007, per request by Peter Wosh.</userDefinedText1>
<subjectLink>
  <subjectTerm>New York City history and culture</subjectTerm>
  <subjectTermType>Topical Term (650)</subjectTermType>
  <subjectSource>Library of Congress Subject Headings</subjectSource>
</subjectLink>
<nameLink>
  <nameLinkFunction>Creator</nameLinkFunction>
  <name>
    <nameType>Corporate Body</nameType>
    <nameSource>local</nameSource>
    <corporatePrimaryName>Public History Program</corporatePrimaryName>
  </name>
</nameLink>
</record>
</accessionRecords>
```