Using ReNamer for Batch Finding Aid Production from Archivists' Toolkit

Cassandra A. Schmitt

University of Maryland

This article will examine the use of the software program ReNamer to facilitate and streamline finding aid production at the University of Oregon. ReNamer was used to perform bulk renaming of EAD files exported from the Archivists' Toolkit (AT) to meet local and consortia file naming schemes. Topics covered include batch exporting from AT, capabilities of ReNamer, and developing workflows to increase productivity.

Background:

AT is an open source archives content management system to support the main functions and tasks performed in an archival repository. AT is designed to help create efficiencies and promote data standardization. In Fall 2013 a new system, ArchivesSpace, launched combining the features of AT with Archon, another archives content management system.

SCUA began using the AT in January 2009 for new accessions. Over the next two years the Archivists' Toolkit Working Group migrated data from multiple sources to AT. In July 2011, older systems went off line and staff started using AT for the full scope of content management functions including accessioning, arrangement and description, production of finding aids and MARC records, and locations management.

Finding Aid Production and NWDA Converter

One of the main benefits and reasons for implementing AT was the capability to export an EAD finding aid from a resource record. However, many repositories will find that the resulting XML file does not meet their standards. The AT working group[i], which also included members from Oregon State University (OSU), needed to ensure that EAD exported from AT would be compatible with the standards, best practices, and requirements of Northwest Digital Archives (NWDA). Both UO and OSU are founding members of NWDA and submit all of our finding aids to the database.

The working group identified areas of the code we wanted changed, developed requirements, and tested a converter script developed by Ryan Wick of OSU. The resulting XSLT style sheet could automatically make a set of changes to an EAD file exported from AT to ensure NWDA compliance. Once this tool was complete, the management was transferred to NWDA, which continues to manage and update as needed and makes the tool available to all NWDA members.

The new process to create an EAD finding aid at SCUA included:

- Entering all relevant information into a resource record in AT following local guidelines
- If short container list, enter directly into AT
- If longer container list, enter into Microsoft Excel template. Convert to XML using Microsoft Word mail merge function. Import EAD file into AT. Merge with main resource record.
- Export resource record from AT into EAD file
- Upload EAD files to NWDA AT EAD Converter tool to automate XML changes to file for NWDA compliance
- Download NWDA compliant EAD file
- Remove "-converted" from file name for each file
- Upload NWDA compliant EAD file to NWDA database

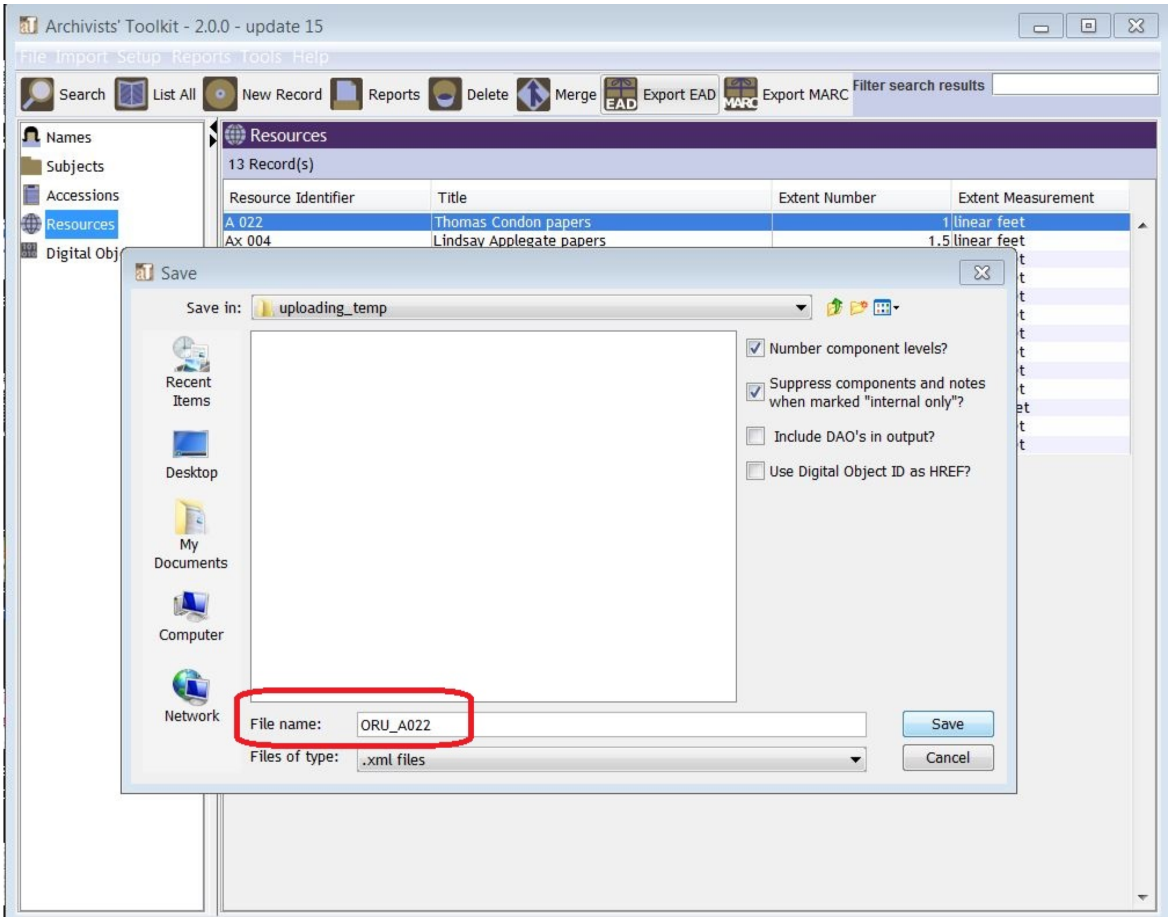
This process, generally, worked well for individual finding aids. After using the NWDA converter "-converted" was added to the end of the file name by the tool so users can identify the new, altered file. The added characters were removed by hand before upload to the database. If we had exported finding aids from AT individually, but were uploading a batch at a time to NWDA, the individual cleanup of the file names added to the length of the process.

Batch Exporting from AT File Names

During the same period, SCUA restructured its processing program and began producing finding aids at a much higher volume than before (in part thanks to a NHPRC basic processing grant). Our goal with these changes, as well as introducing AT, was to create efficiencies.

In addition to the increasing amount of time it took to remove “-converted” from the end of each file, we discovered a quirk in AT when batch-exporting EAD records that amplified the problem of hand editing file names.

When you individually export a single finding aid from AT, the program prompts you to enter the file name into the save dialog box. The system prompts you to do this whether you are in the resource record and press the “Export EAD” button or if you are a list view, have the record highlighted, and press the “Export EAD” button.



13 Record(s)

Resource Identifier	Title	Extent Number	Extent Measurement
A 022	Thomas Condon papers	1	linear feet
Ax 004	Lindsay Applegate papers	1.5	linear feet

Save in: uploading_temp

☑ Number component levels?

☑ Suppress components and notes when marked "internal only"?

☐ Include DAO's in output?

☐ Use Digital Object ID as HREF?

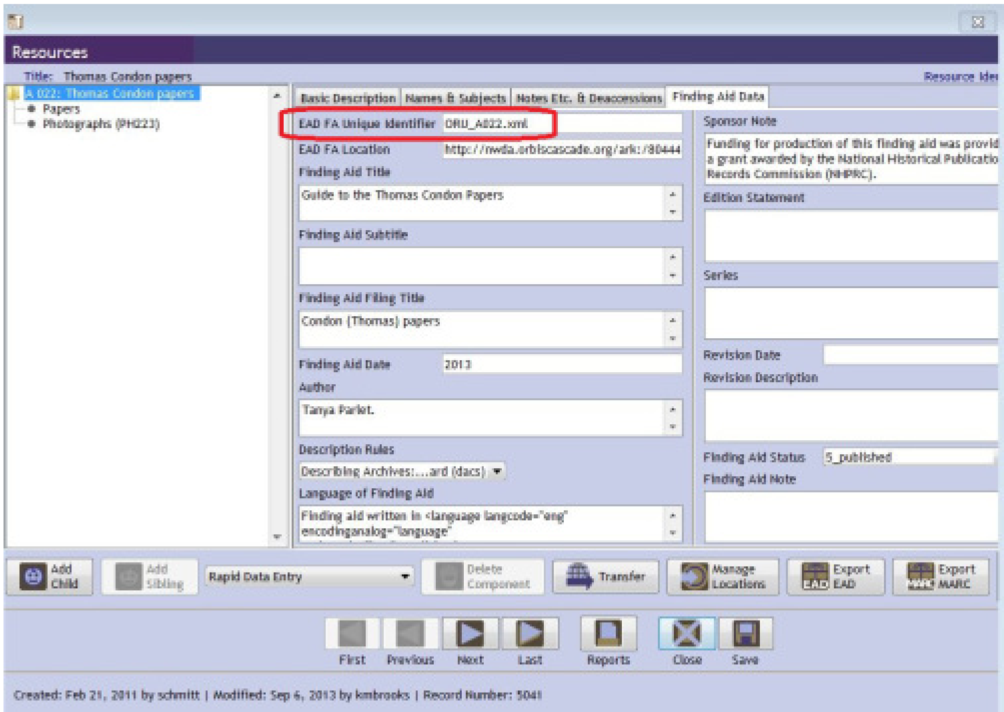
File name: ORU_A022

Files of type: .xml files

Figure 1: Exporting individual EAD files allows you to supply a title during the process

When batch exporting finding aids, AT automatically names the files instead of the user supplying the name into the save dialog box. This makes sense as the bulk option is meant to expedite the process and could involve several to a few hundred files at a time.

However, the quirk arrives in how AT names these files during the bulk export. Within AT on the Finding Aid Data tab, there is a field, EAD FA Unique Identifier, which acts as the unique identifier for the finding aid XML file. Most repositories will follow a naming convention for this field. All our resource records have this field filled out before any export following local and NWDA guidelines.



Title: Thomas Condon papers

A 022: Thomas Condon papers
- Papers
- Photographs (PH223)

EAD FA Unique Identifier: ORU_A022.xml

EAD FA Location: http://nwda.orbiscascade.org/ark:/80444

Finding Aid Title: Guide to the Thomas Condon Papers

Finding Aid Subtitle:

Finding Aid Filing Title: Condon (Thomas) papers

Finding Aid Date: 2013

Author: Tanya Parlet.

Description Rules: Describing Archives:...ard (dacs)

Language of Finding Aid: Finding aid written in <language langcode="eng" encodinganalog="language"

Sponsor Note: Funding for production of this finding aid was provid a grant awarded by the National Historical Publicatio Records Commission (NHPRC).

Edition Statement:

Series:

Revision Date:

Revision Description:

Finding Aid Status: 5_published

Finding Aid Note:

Created: Feb 21, 2011 by schmitt | Modified: Sep 6, 2013 by kmbrooks | Record Number: 5041

Figure 2: AT field EAD FA Unique Identifier

One would assume that AT would use the value of this field to name EAD files as this is the purpose of the field. Instead AT combines the Resource Identifier fields and the text "-ead" to create the file name.

Upon discovery of this issue, we clarified our workflow:

- Export resource record from AT into EAD file
  - If individual, type in file name using EAD FA Unique Identifier value during export/save process
  - If part of a batch export, correct file name for each file after export using EAD FA Unique Identifier value

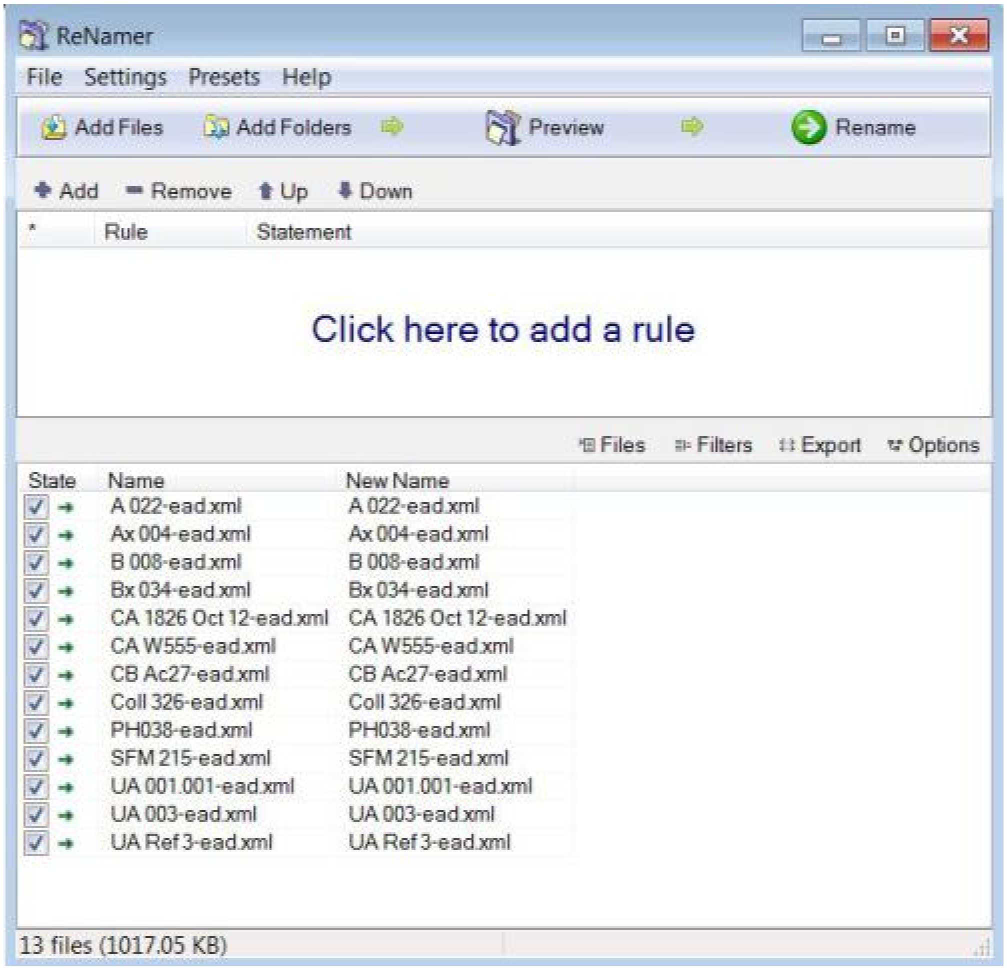
The bulk-naming scheme resulted in additional timely hand editing of file names for batch produced (or updated) EAD finding aids. This was in addition to the file name manipulation for every finding aid after using the NWDA AT EAD conversion tool. While this process was still an improvement in efficiency from our pre-AT days, the two sets of hand change to file names did not bring us to the level of efficiency we desired.

Automating File Name Changes

Under these circumstances, we started looking for tools that would help automate the file name correction processes. There is a plethora of file renaming tools available for both Windows and Mac operating environments. However, most tools seem to be geared towards one or the other and not both. We reviewed the tool descriptions and read reviews from users and software business sites.

After performing some research into the functionalities of different tools, we decided to test the free software program ReNamer, which allows you to perform batch file renaming operations. The tool was developed by Denis Kozlov, a software developer currently based out of Dublin, Ireland. The software is free to download with the option of purchasing a commercial license. Full documentation is available on den4b wiki. (Confession: I didn’t use any of the documentation, just played around and figured it out.) ReNamer is only available on the Windows platform, which worked for our office setup.

Integrating ReNamer into our workflow was pretty straightforward. After exporting EAD finding aids in a bulk operation, files were put in a staging area folder on the server. This folder was then reloaded into ReNamer by clicking the “add folders” button. ReNamer has the option of performing changes to a group of files or a group of folders. If you add folders to the tool, changes will be made to all the files within those folders, so pick carefully or make sure your folder structure enables the changes you want.



State	Name	New Name
☑ →	A 022-ead.xml	A 022-ead.xml
☑ →	Ax 004-ead.xml	Ax 004-ead.xml
☑ →	B 008-ead.xml	B 008-ead.xml
☑ →	Bx 034-ead.xml	Bx 034-ead.xml
☑ →	CA 1826 Oct 12-ead.xml	CA 1826 Oct 12-ead.xml
☑ →	CA W555-ead.xml	CA W555-ead.xml
☑ →	CB Ac27-ead.xml	CB Ac27-ead.xml
☑ →	Coll 326-ead.xml	Coll 326-ead.xml
☑ →	PH038-ead.xml	PH038-ead.xml
☑ →	SFM 215-ead.xml	SFM 215-ead.xml
☑ →	UA 001.001-ead.xml	UA 001.001-ead.xml
☑ →	UA 003-ead.xml	UA 003-ead.xml
☑ →	UA Ref 3-ead.xml	UA Ref 3-ead.xml

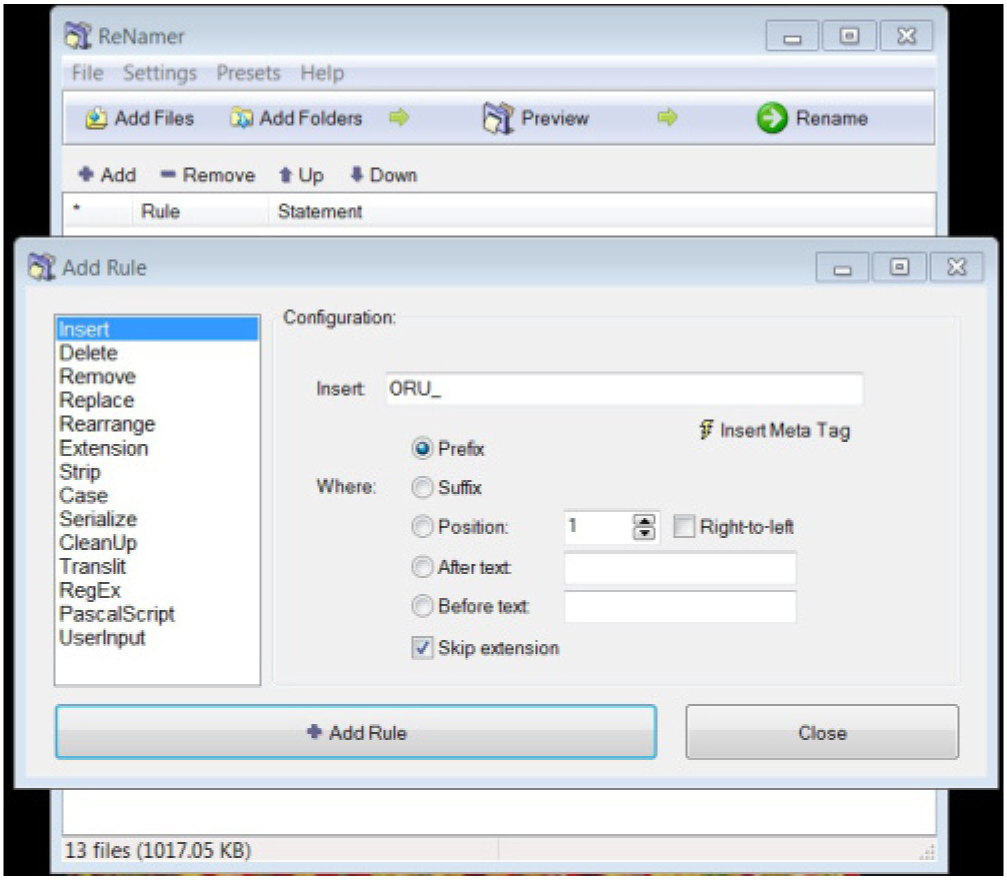
13 files (1017.05 KB)

Figure 3: Folder of exported EAD files in ReNamer

I recommend adding your files to the program before adding any rules, as the program will preview the changes. This will allow you to alter or change any rules or actions before executing the name changes.

ReNamer is extremely flexible and allows you to change file or folder names in multiple, complex ways. You can add or remove characters based on a number of factors including insert, delete, remove, replace, rearrange, extension, strip, case, serialize, cleanup, translit, regex, pascalscript, userinput. Each of these then has multiple options. For example, for the insert rule you can insert characters to a prefix, suffix, position number, after text, before text, and an option of skipping the extension when you want.

Some of the rules have the capability to utilize “MetaTags”. This includes data embedded in the file such as the title, size, creator, version information, etc…



Insert: ORU_

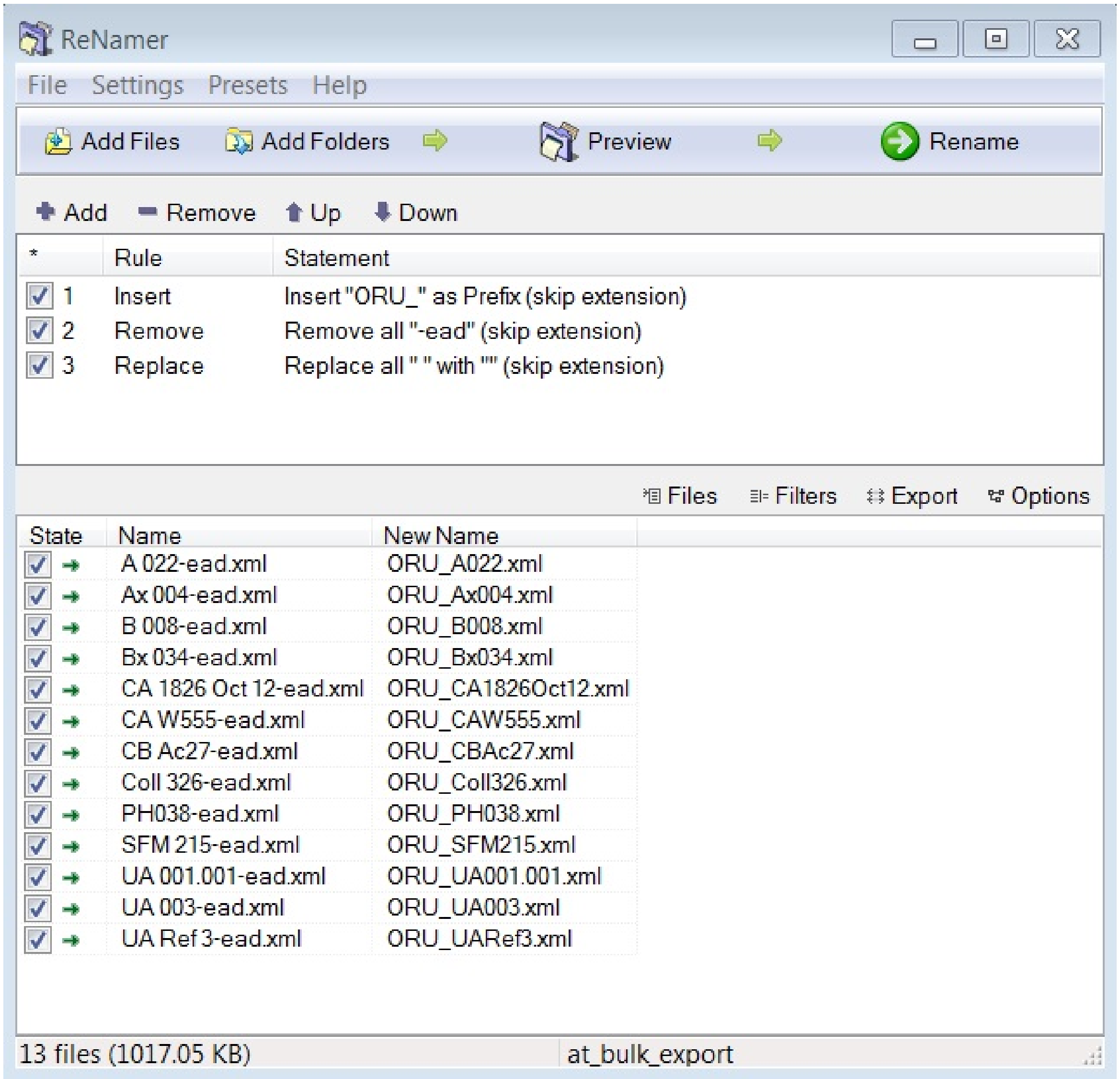
Where:
● Prefix
○ Suffix
○ Position: 1 □ Right-to-left
○ After text
○ Before text
☑ Skip extension

13 files (1017.05 KB)

Figure 4: Insert rule in ReNamer

To make the file names match our requirements and the values in the EAD FA Unique Identifier field, we added three rules performing different operations:

- Inserted “ORU_” to the beginning of the file name
- Removed “-ead” from the end of the file name
- Removed the space from the resource identifier portion of the file name



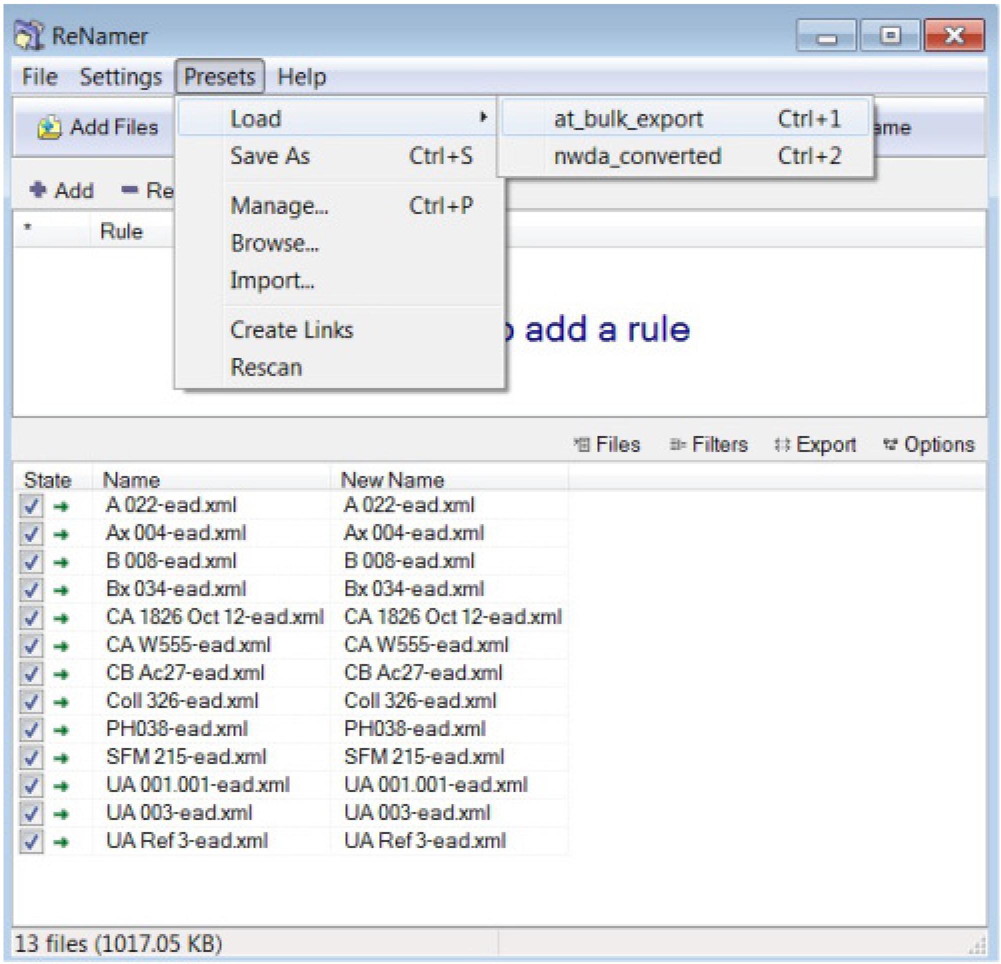
*	Rule	Statement
☑ 1	Insert	Insert "ORU_" as Prefix (skip extension)
☑ 2	Remove	Remove all "-ead" (skip extension)
☑ 3	Replace	Replace all " " with "" (skip extension)

State	Name	New Name
☑ ➜	A 022-ead.xml	ORU_A022.xml
☑ ➜	Ax 004-ead.xml	ORU_Ax004.xml
☑ ➜	B 008-ead.xml	ORU_B008.xml
☑ ➜	Bx 034-ead.xml	ORU_Bx034.xml
☑ ➜	CA 1826 Oct 12-ead.xml	ORU_CA1826Oct12.xml
☑ ➜	CA W555-ead.xml	ORU_CAW555.xml
☑ ➜	CB Ac27-ead.xml	ORU_CBAc27.xml
☑ ➜	Coll 326-ead.xml	ORU_Coll326.xml
☑ ➜	PH038-ead.xml	ORU_PH038.xml
☑ ➜	SFM 215-ead.xml	ORU_SFM215.xml
☑ ➜	UA 001.001-ead.xml	ORU_UA001.001.xml
☑ ➜	UA 003-ead.xml	ORU_UA003.xml
☑ ➜	UA Ref 3-ead.xml	ORU_UARef3.xml

13 files (1017.05 KB) at_bulk_export

Figure 5: Previewing file name changes using a custom preset

ReNamer also allows you to set “presets”, a group of changes to be performed all at once. If you have a set of changes you would like to make on a regular basis, you can save the group as a preset. You will then be able to execute them all at once instead of creating and executing each change individually or having to recreate and add each rule every time.



State	Name	New Name
☑ →	A 022-ead.xml	A 022-ead.xml
☑ →	Ax 004-ead.xml	Ax 004-ead.xml
☑ →	B 008-ead.xml	B 008-ead.xml
☑ →	Bx 034-ead.xml	Bx 034-ead.xml
☑ →	CA 1826 Oct 12-ead.xml	CA 1826 Oct 12-ead.xml
☑ →	CA W555-ead.xml	CA W555-ead.xml
☑ →	CB Ac27-ead.xml	CB Ac27-ead.xml
☑ →	Coll 326-ead.xml	Coll 326-ead.xml
☑ →	PH038-ead.xml	PH038-ead.xml
☑ →	SFM 215-ead.xml	SFM 215-ead.xml
☑ →	UA 001.001-ead.xml	UA 001.001-ead.xml
☑ →	UA 003-ead.xml	UA 003-ead.xml
☑ →	UA Ref 3-ead.xml	UA Ref 3-ead.xml

13 files (1017.05 KB)

Figure 6: Loading saved presets in ReNamer into current project

After previewing files, we use the rename button to perform the operations. ReNamer will display the changes and confirm or deny success. We then proceed using the NWDA AT EAD converter and download the converted files with “-converted” appended.

To remove the “-converted” from the file name we opened a new project in ReNamer, loaded our folder with the NWDA converted files, and used another preset to make our final changes.

Conclusion

While we had to add extra steps and a new tool to our process, introducing ReNamer eliminated individual editing of file names. For an individual record, this new process did not save time, but was immensely helpful for batch operations. Testing rules and setting up the presets was a quick process. As we increased our production rate of finding aids the procedures and tool was scalable and our benefit from the time savings continued to grow.

As our local best practices change and we utilize new tools for finding aid production, we can continue to elevate the use of ReNamer and will make adjustments as necessary. The implementation of ReNamer also provided a successful example of using new (to us) software and technologies to the department. It demonstrated that some technology changes can happen relatively quickly with low barriers to entry and development.

About the Author

Cassandra A. Schmitt is the Head of Access for Special Collections and University Archives at the University of Maryland, College Park where she oversees technical service activities. Previously she was the Archivist for Collections Management at the University of Oregon where she managed accessioning, arrangement and description, finding aid production, management of Archivists' Toolkit, and other technical projects. From 2008-2011 she was the Accessioning and Processing Archivist at UO. She has been active in both regional and national organizations, particularly around issues involving EAD and the Archivists' Toolkit. Cassie has an MSI from the University of Michigan where she worked at the Bentley Historical Library, Clements Library, and Special Collections Library and a BA in History and Philosophy, Politics, and Law from Binghamton University.

Notes

[i] Members of the Archivists’ Toolkit Working Group involved in the converter script development included: Larry Landis, Elizabeth Nielsen, and Ryan Wick from Oregon State University; Nathan Georgitis and Cassandra A. Schmitt from the University of Oregon.