

Molly Schwartz & Ruth Kitchin Tillman

MARAC 2012 Fall Conference

Embracing Archivists' Toolkit to Implement EAD

Improving the quality and standardization of metadata across the archival field would provide obvious benefits. Techniques for attaining consistent and uniform metadata usage have been the subject of articles, presentations, workshops, projects, and grants. And yet, even as archivists and programmers devise metadata standards specifically designed for archival collections and the software to implement them, studies have shown that many repositories are not taking advantage of these tools. In a 2010 article by Christopher J. Prom in *The American Archivist*, he analyzes a study conducted by Mark Greene and Dennis Meissner on current processing practices in college and university archives. Prom finds that, despite the archival profession's promotion of Encoded Archival Description (EAD) for salutary effects, "such as allowing for interoperability, encouraging good descriptive habits, and standardizing presentation,"¹ of the archives surveyed only "9% see EAD as the principal format at the folder level"² because "it is currently beyond the capacity of many institutions to implement...EAD in a cost-effective fashion."³

Given these statistics, we undertook this project, which is a combination of literature analysis and hands-on experimentation, to develop a strategy by which aspiring, inexperienced, and less tech-savvy archivists can implement EAD and create finding aids with high quality metadata using available software, specifically Archivists' Toolkit. This

¹ Prom, Christopher J. "Optimum Access? Processing in College and University Archives," *American Archivist* 73:1 (Spring/Summer 2010), 148.

² *Ibid.*, 161.

³ *Ibid.*, 173.

project, in which a group of archives graduate students with no prior experience with EAD used Archivists Toolkit to successfully describe and export a collection in EAD, demonstrates that all repositories can adopt EAD by using software such as Archivists' Toolkit. According to a second article by Christopher Prom, the fact that smaller college and university archives are trailing behind large institutions in the adoption of EAD is to be expected because "implementing a new technology requires both rigorous training and a funding source."⁴ When Prom wrote this article in 2002, he called for advances in authoring and browsing software that improves EAD's ease of use for archivists, is more archivist friendly, and "presents finding aids in a fashion that is understandable and helpful to archival users."⁵ Today, we have access to this type of software. We have Archivists' Toolkit and Archon, and in the next few years the best features of each will be combined with the creation of ArchivesSpace. As graduate students, we are operating from the perspective that it is especially crucial that we, as the next generation of archivists, enter our jobs with the training and knowledge to arrange and describe records using EAD and use software that facilitates EAD, such as Archivists' Toolkit.

Intro to EAD

To review, Encoded Archival Description (EAD) was developed in the 1990s as part of a project initiated by the University of California, Berkeley Library. It is a non-proprietary encoding standard for metadata about an archival collection which presents the relationships and hierarchies within the collection. The developers initially based EAD on the Standard Generalized Markup Language (SGML), but decided shortly before its release that the newly-created eXtensible Markup Language (XML), also based on

⁴ Christopher Prom, "The EAD Cookbook: A survey and Usability Study," *The American Archivist* 65, no. 2 (2002): 262, Accessed June 9, 2012, <http://www.jstor.org/stable/40294209>

⁵ *Ibid.*, 258-259.

SGML, was better suited to be the parent language.⁶ As an XML-based language, EAD can be transformed into X/HTML, plain text, and PDF, using eXtensible Stylesheet Language (XSL). EAD is currently the preferred machine-readable descriptive system for finding aids.

This paper will primarily discuss Archivists' Toolkit, one of whose main strengths is that it allows users to describe a record and export it as EAD. Archivists' Toolkit also supports the standardization and interoperability of EAD metadata because it discourages local repositories from using localized models that hamper interoperability⁷ and ensures tighter best practices by requiring that EAD files includes key metadata elements and use standard coding patterns.⁸ While those responsible for displaying finding aids and other collection-related information on archival websites will need a detailed understanding of EAD, Archivists' Toolkit allows any archivist to create well-formed EAD records without learning anything about the language.

Archivists' Toolkit

Developed and maintained with grants from the Andrew W. Mellon Foundation, the Archivists' Toolkit is an open-source database application designed to facilitate the processing and management of archival information in every step of the life-cycle of a record, including accessioning; registering source information; providing access points for subjects and creators of resources; tracking locations of archival resources; and producing finding aids, collection guides, and administrative reports.^{9 10} Archivists'

⁶ "Development of the Encoded Archival Description DTD," *Library of Congress*, accessed August 30, 2012, <http://www.loc.gov/ead/eaddev.html>.

⁷ Prom (2010): 150.

⁸ Christopher J. Prom, "Does EAD Play Well with Other Metadata Standards?" *Journal of Archival Organization* 1, no. 3 (2002): 51.

⁹ "Archivists Toolkit to be Open Source." *Library Journal* 129, no. 16 (2004). DOI: 03630277.

Toolkit was designed to be technologically and financially accessible to all archivists and to operate in accord with archival community standards such as ISAD-G, ISAAR(CPF), DACS, EAD, METS, MARCXML, MODS, and Dublin Core.¹¹ Its main strengths lie in that it supports the production of access instruments, promotes metadata standardization, increases efficiency, and lowers training costs.¹² Archivists' Toolkit can be used immediately with little professional training or installation costs, it can be customized with a style sheet, it creates PDF and EAD finding aids, and it preserves the hierarchical aspect of data management.

Advancements

Over time, Archivists' Toolkit has evolved to incorporate changes and advances in technology and the archives community. With additional funding from the Andrew Mellon Foundation, the New York University Libraries, UC San Diego Libraries, and the University of Illinois Urbana-Champaign Libraries are partnering to develop ArchivesSpace, an archives management application that will combine the best features of the Archivists' Toolkit and Archon.¹³ This integration of the Archivists' Toolkit and Archon is set to be rolled out in May 2013 and is expected to “result in a functionally superior archives management / access tool for AT and Archon users alike [...] that will be easier to sustain over time.”¹⁴ The creation of ArchivesSpace will resolve one of the main criticisms of Archivists' Toolkit, which is that it does not support web-publishing of

¹⁰ “Introduction to the Archivists' Toolkit,” *Archivists' Toolkit*, accessed September 9, 2012, <http://archiviststoolkit.org/sites/default/files/Chapter%201.pdf>

¹¹ Ibid.

¹² Holly Mengel, “The Archivists' Toolkit,” Presentation at the MARAC Fall 2010 Conference, Harrisburg, PA, November 11-13, 2010, <http://www.slideshare.net/hollymengel/at-marac>

¹³ Archon is another archives management software, admired for its web publishing features.

¹⁴ Theimer and Westbrook.

finding aids.¹⁵ The ArchivesSpace team will be performing extensive testing before release, ensuring that the product will meet the needs of the archives community and will release without major design flaws. By combining the strengths of the two leading archives management applications, Archivists' Toolkit and Archon, the ArchivesSpace team will be providing the archives community with an updated, relevant, and efficient method for creating EAD finding aids for all collections. Archivists' Toolkit will essentially be upgraded to incorporate the latest technologies, meaning that all archivists with skills and experience using Archivists' Toolkit will be able to transfer these skills to the new ArchivesSpace software. When released, ArchivesSpace will also allow repositories using Archivists' Toolkit to import their entire databases.

Feedback/Implementation

Many repositories have chosen to implement Archivists' Toolkit due to its ease of use, strong support system, and its unique suitability to archives. Repositories that adopt Archivists' Toolkit do so in part because there are support systems ensuring that archivists new to the software will be able to learn how to use it, someone will know how to fix it if something goes wrong, and that the program will not disappear in the next few years. The Mellon Foundation has provided consistent and ongoing grant support for Archivists' Toolkit, and online resources such as the Society of American Archivists (SAA) AT and Archon Round Table and the AT Wiki publish guides about how to use the program, forums for sharing issues encountered, and news bulletins about updates to the software.

¹⁵ Ibid.

In 2008, the “archivalsoftware” wiki posted excerpts of interviews conducted with five archivists about how they use Archivists’ Toolkit. Although this feedback does not reflect the changes made to Archivists’ Toolkit in its most recent versions and it is a very limited sample, it still conveys some important points about why archivists chose to use Archivists’ Toolkit, how easy it was to install and use, and its strengths and weaknesses. One archivist who had experience teaching two Archivists’ Toolkit workshops observed that, although people learned how to use it pretty quickly and it was easy for archivists to use, “someone who wasn’t trained as an archivist had some problems with it; it’s set up with the assumption that you are an archivist.”¹⁶ Another archivist noted that, while it was simple to teach students and staff how to input information into Archivists’ Toolkit, “there needs to be someone in the department with a more thorough understanding of the program and how things work... -- otherwise you will run into databases that are not very standardized.”¹⁷ A third archivist commented that, due to the “loosey-goosey” nature of many of the notes sections that do not use controlled vocabularies, “it will be important for each repository to do the intellectual work up front of giving grad students good guidelines about how to formulate data.”¹⁸ The general consensus of all five archivists using the software was that while Archivists’ Toolkit is easy for anyone, regardless of technological or archival background, to input data into, being able to use the software effectively requires a basic knowledge of archival principles and a repository-specific policy and set of guidelines. This study seeks to clarify how successfully inexperienced

¹⁶ Lisa Spiro, “Archivists’ Toolkit User Feedback,” *archivalsoftware*, November 23, 2008, <http://archivalsoftware.pbworks.com/w/page/13600239/Archivists%20Toolkit%20User%20Feedback>.

¹⁷ Ibid.

¹⁸ Ibid.

graduate students in an archives program can train themselves to understand and use AT and how they can use this knowledge in their future careers.

Our Project

For the term project of a University of Maryland Organization of Information class, a group of archives students chose to explore Archivists' Toolkit and examine how it interacted with EAD files created by other methods, with the goals of preparing themselves to work in archives which use the software and to be aware of the factors which would need consideration at a repository transitioning its EAD files into Archivists' Toolkit. Thanks to the help of Jennie Anne Levine Knies, manager of Digital Collections at the University of Maryland, the group used one of the university's EAD files.¹⁹ ²⁰ The project consisted of: 1) importing the UMD finding aid into AT, 2) exporting the finding aid from the resulting file and comparing the two, 3) and individually recreating the finding aid based on the UMD file and web display, then comparing these records within the group and to the original file. This presentation is concerned with the first and second steps, although it will touch lightly on the third. In preparation for this presentation, one of the authors imported a second non-Archivists' Toolkit EAD record from another repository to compare import experiences and results. While the technical examination of the EAD import process will concern itself primarily with the import of the UMD file, the second will occasionally be mentioned, as it provided several useful examples of problems during import. This project dealt with the

¹⁹ The EAD creation method used by the University of Maryland can be found in Ms. Knies's article "Taming the 'Beast': An Archival Management System Based on EAD." *Journal of Archival Organization*, v.4, issue 3-4 (2007).

²⁰ <http://digital.lib.umd.edu/oclc/MdU.ead.histms.0001.xml>

records aspect of Archivists' Toolkit, not its accessions module, digital objects, or names and subjects.

The parent-level fields which Archivists Toolkit requires in order to save any record which is being created or edited for the first time are:

1. Resource Identifier,
2. Level,
3. Title,
4. Date,
5. Language Code,
6. Extent,
7. and Extent Type.²¹

Despite not containing all of the seven fields required by Archivists' Toolkit, both records imported successfully. After each import, the import log specified which fields must be added to validate the record in its system. The record's invalid status did not prevent ingest or viewing, but all the required fields had to be filled out before the record could be saved after editing. While the lack of certain fields did not prevent import, the file from the second repository could not initially be imported because it contained a tag, which is not a part of the EAD DTD or EAD Schema.²² In this case, the tag was being used to designate a title, so the <title> tag was substituted in the EAD file and the

²¹ For child levels, it requires the Level and either a Title or Date field to be filled out.

²² An alphabetical list can be found here: http://www.loc.gov/ead/tglib/element_index.html

import succeeded.²³ Thus the import experiment demonstrated that any record which conforms to EAD can be imported into Archivist's Toolkit, whether or not it contains all the elements Archivist's Toolkit requires to create or edit a record, but that records which use elements outside the EAD DTD or Schema will fail to import.²⁴

Learning to Use Archivist's Toolkit

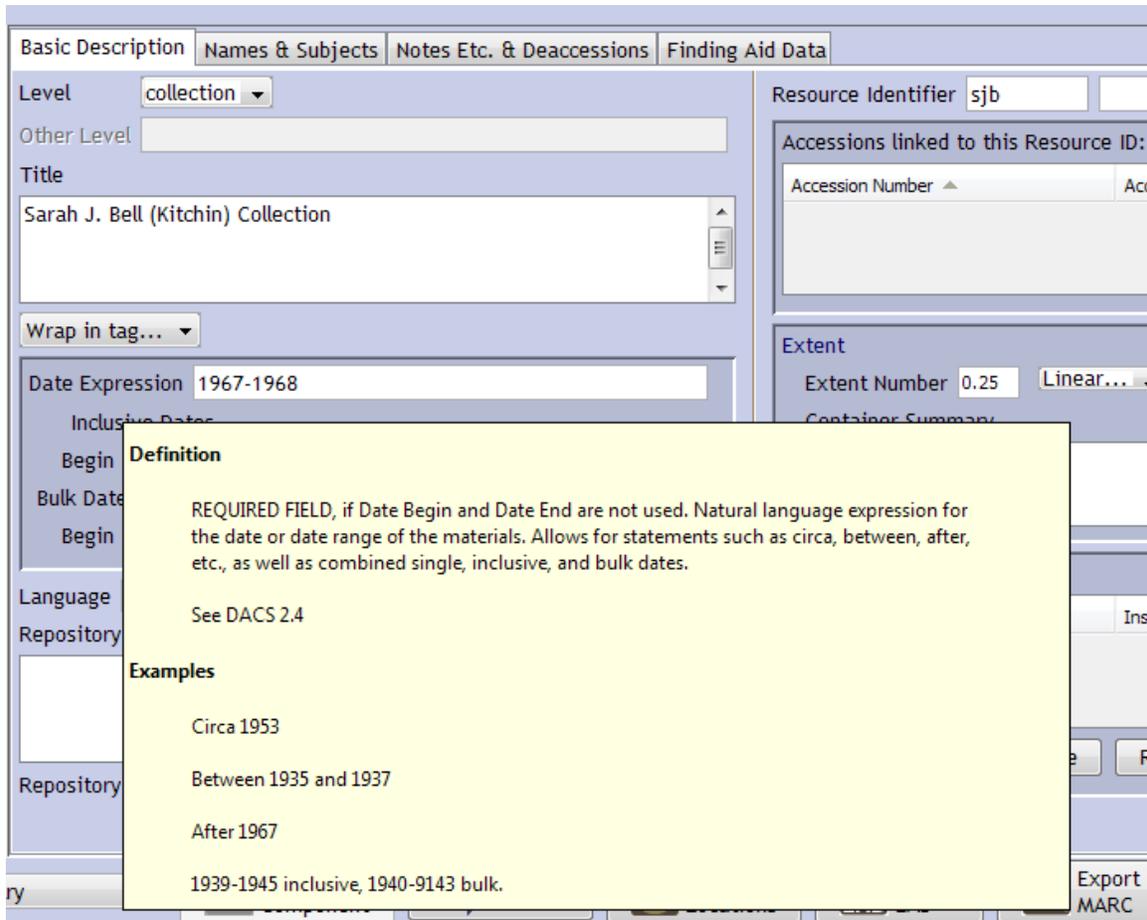
Even though the project team was mostly unfamiliar with XML and fairly new to archival description, each member figured out individually how to use the software. As part of the exploratory process, each contacted the others as little as possible at first, so that different learning experiences could be discussed afterward. Rather than installing and initializing databases on each computer using the instructions on Archivist's Toolkit's website, the team used the Archivists' Toolkit Sandbox database to store its records. This made the initial setup far simpler and allowed each member to see the others' work and to help answer questions.

Archivist's Toolkit is designed to help even the less tech-savvy figure out what its numerous fields mean. New users who are confused about fields can simply hold the mouse over a field's name and read a short description of the field, its citation in DACS, and some examples (as pictured below). The one way in which they could improve this feature would be if they added an example of which EAD element corresponded to the entry, as sometimes it's less intuitive than something like <bioghist> for Biographical

²³ The element <emph> could also have been used if the archivist wanted to emphasize something in the record. Both <title> and <emph> can be displayed as bold or italic using CSS or can have attributes such as @render=bold or @render=italic within the tag itself to determine a specific display.

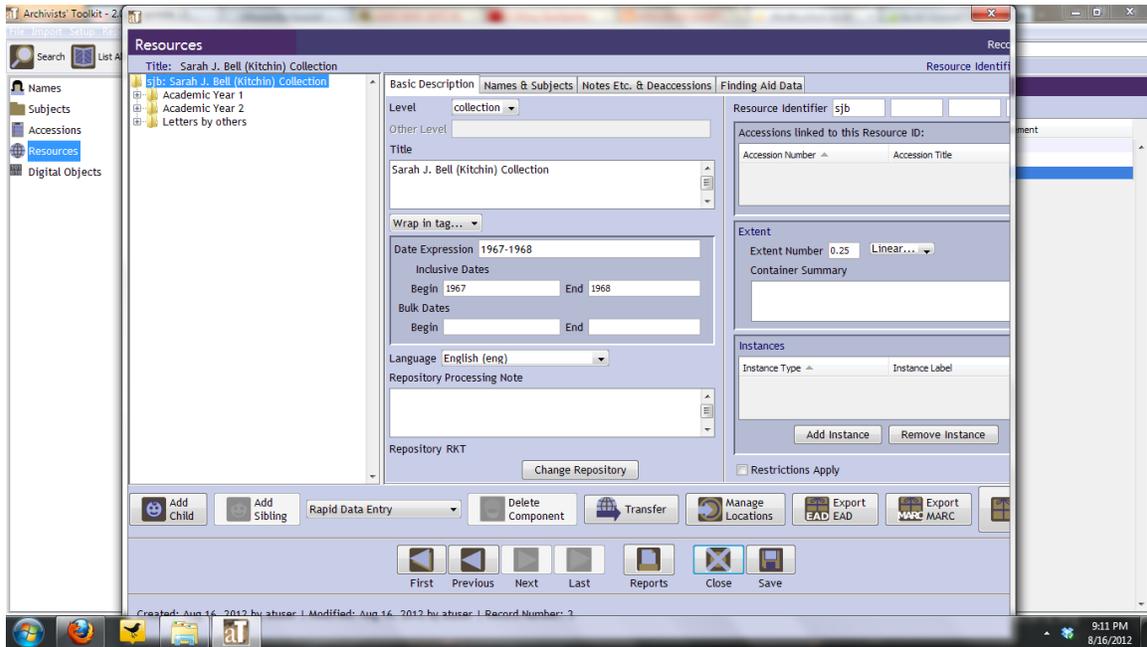
²⁴ According to the handbook for versions 1.5 and 1.59, the most recent one available since the handbook for 2.0 is still being written, when Archivist's Toolkit does batch imports, it will judge each file on its own merits, so one improper file will not stop the import, nor will an incomplete file http://archiviststoolkit.org/sites/default/files/AT1_5_UserManual.pdf p.67

and Historical note. The Archivists' Toolkit website does not have any apparent documentation of field/element correspondence.

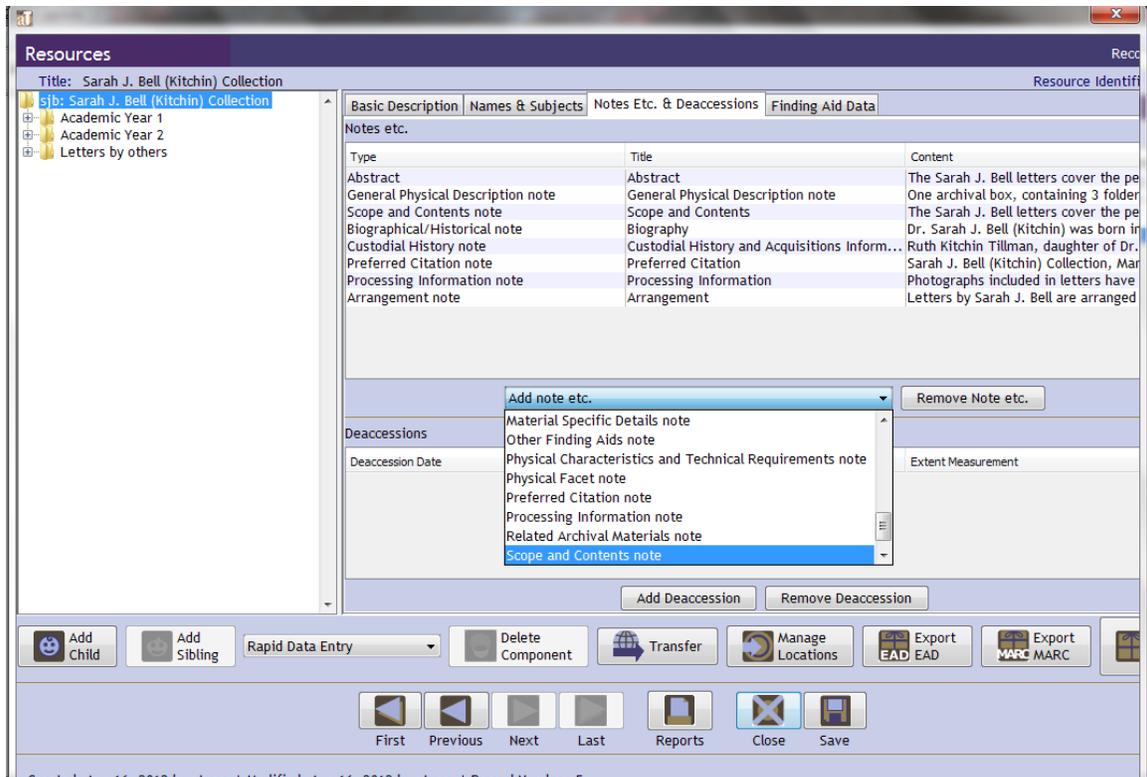


Creating a basic record was quite straightforward. It does not require the user to know anything about EAD, one just has to understand the basics of archival theory and description—such as what should go in a “Scope and Contents note.” Additionally, one should understand the repository’s policies on levels of description and necessary contents for the record. Repository managers may aid the process by creating specialized templates for rapid data entry which only include the fields required for certain collections and may include preset information.

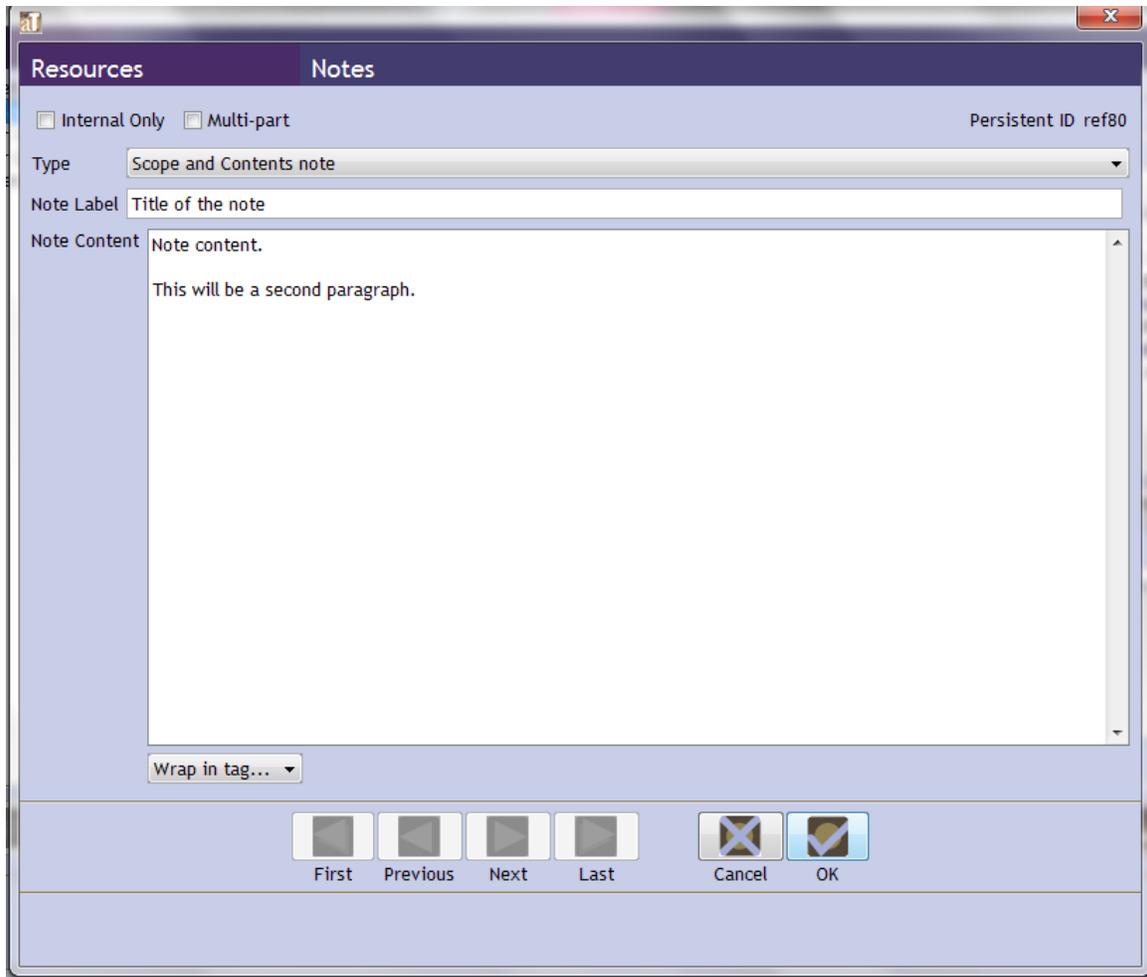
The record creation and editing interface is divided into four clearly-named sections: “Basic Description,” “Names & Subjects,” “Notes Etc. & Deaccessions,” and “Finding Aid Data.”



Adding normal fields to the Scope and Content note, for example, is quite simple. One goes into “Notes Etc. & Deaccessions,” selects “Scope and Contents note” from the drop-down menu,



and then enters the note's label (to be used as a title), and the note's content.



The note creation popup includes the options to make it internal and/or multi-part, and to select text to wrap in a tag. If the tag can have attributes, one may specify them, but one needn't. Paragraph breaks will be parsed into separate `<p>` tags during export.

The screenshot shows both the simplicity and the biggest drawback of AT. While the user interface is easy-to-learn, it is also ugly and clunky. Maximizing some windows isn't possible without manually dragging the sides. Some information may be partially off-screen unless the window is maximized and there is no horizontal scroll-bar to allow one to navigate over without maximization. Linux and Mac systems automatically create an option to maximize and then fit the window to the screen.

Team members experienced an initial learning curve, followed by a basic understanding of the software. However, upon comparing records within the group, it was discovered that while they all contained essentially the same information, the information was structured within elements that varied from member to member.²⁵ This result reinforces the archivist's statement earlier that departments must do a certain amount of preparatory work in order to get standardized records with proper authority control.

Quality of Metadata Exported

When the exported files were tested against the EAD Schema, they were found to contain no errors. While it cannot control quality of metadata entered by an archivist or graduate assistant, Archivists' Toolkit has been carefully designed to work within the EAD Schema and exports properly. The EAD export function allows users to designate whether or not they want components to export inside numbered or unnumbered elements, whether fields marked as internal should be exported, whether to include digital objects, and how to handle digital object IDs. As a bonus to users, the export file was indented in a manner which made it easy for one to visually parse the nested data.

Strategy

As this experiment demonstrates, graduate students in library science with no prior experience arranging and describing records have the capability and resources to create EAD style sheets for archival collections using Archivists' Toolkit. With some additional knowledge of XML, HTML, and XPath, graduate students can convert these style sheets into online finding aids. In the current competitive job market, it is important

²⁵ The team was working from the record as displayed on the University of Maryland's website rather than from the EAD file itself, in order to better simulate the experience of creating a record using data collected offline.

that archives graduate students acquire practical skills while they are in school. Graduate students must take advantage of software such as Archivists' Toolkit that are easily accessible, facilitate the creation of finding aids with quality metadata, and do not require advanced cataloging or technological skills. The following is a strategy for graduate students to teach themselves how to use Archivists' Toolkit and put this skill on their résumés.

Archival Principles

In order to truly understand the hierarchical relationships between records and the descriptive elements, users of Archivists' Toolkit should be familiar with the fundamental principles of archives. As the byproducts of human daily activity, records can come in a wide variety of types and materials over which archivists must establish physical and intellectual control.²⁶ This influences the descriptive elements in EAD and Archivists' Toolkit and necessitates that many descriptive fields in Archivists' Toolkit do not use controlled vocabularies. It is also important to understand that archivists organize at the group rather than the item level and that these groups are organized hierarchically. The metadata generated by Archivists' Toolkit reflects the archival emphasis on context over content when describing materials. These archival principles played a large role in the creation of Archivists' Toolkit, and therefore it is useful for graduate students to be familiar these principles in order to understand how to use Archivists' Toolkit.

Basics of EAD

Although Archivists' Toolkit facilitates the creation of EAD finding aids, it is useful for Archivists' Toolkit users to have a basic knowledge of EAD if they are using

²⁶ Mary Jo Pugh, *Providing Reference Services for Archives and Manuscripts*, Chicago: Society of American Archivists, 2005, 13.

EAD as the metadata standard for the organization of their materials. Handbooks and explanations of the EAD metadata schema are freely available on the web, and most of them are very thorough and written in laymen's terms. Because Archivists' Toolkit imports and exports descriptions in EAD, AT users do not need to know how to create descriptions in EAD. It is important, however, to be able to verify the quality of the metadata that Archivists' Toolkit imports and fix mistakes or inconsistencies and ensure that exported files meet the repository's internal requirements.

Practice in the Sandbox

After reviewing archival principles and fundamentals of EAD, graduate students are ready to start learning how to use Archivists' Toolkit in a hands-on fashion via the Archivists' Toolkit Sandbox. The Archivists' Toolkit provides a "sandbox" on their website, which allows anyone to download Archivists' Toolkit and sign in with a test user name and password, and play around with the software.²⁷ An excellent way for students to learn how to use Archivists' Toolkit is to repeat the procedures of this study: access the XML files of EAD finding aids of real collections, practice importing the record to Archivists' Toolkit, manually create a finding aid by inputting all the information by hand, export the collection as an XML file, and evaluate the result.²⁸ Archivists' Toolkit provides a tremendous opportunity to graduate students to develop valuable, practical cataloging skills. With some foresight, proactive initiative, and practice, all graduate students can put Archivists' Toolkit on their résumés when they apply for jobs.

²⁷ Sandbox info can be found here: <http://archiviststoolkit.org/support>

²⁸ When using any Sandbox, practice straightforward courtesy. Do not put anyone's private information in the sandbox, as others may be able to view it. If you are looking at someone else's record in the sandbox, never ever make or save changes. Once you're done with a record on the sandbox, consider deleting it to free up space. Sandbox records may be exported as EAD and imported into one's own database later.

Appendix

In comparing the University of Maryland's EAD file for the Richard White collection²⁹ and the Archivist's Toolkit record created from import along with its export file, several problems were noticed which are not germane to the paper's topic, but which may be of interest to those who are learning Archivist's Toolkit in order to manage EAD records already in existence. These issues are outlined in this appendix.

A. *Multiple date fields*

- 1) The program only imported the first date from the <publicationstmt> field.
- 2) The program imported *both* the @type=inclusive and @type=bulk fields from <unitdate> but conflated them into the same field, yielding an odd-looking "1905-1920 1905-1920."

B. *MARC fields and @encodinganalog*

Archivist's Toolkit did not import MARC fields from @encodinganalog, [nor does it have the option to include the encodinganalog attribute in its EAD export files. The MARC XML export however, seemed to properly interpret the data and corresponded with the encodinganalog as specified in the imported file.

C. *Folder hierarchy*

The folders' @parent field did not import, although their folder type, number, and nesting structure did.

D. *Manipulated data*

²⁹ <http://digital.lib.umd.edu/oclc/MdU.ead.histms.0001.xml>

- 1) In its EAD export file, Archivist Toolkit adds the text “Finding aid processed by” directly after its <author> tag. This will cause problems for any imported file which included more than the author(s) name(s) in that field. [fig.]
- 2) Archivist’s Toolkit removes double-spacing after periods, in deference to current rules of style and, more importantly, XML rules for data not designated as CDATA.
- 3) For note fields which did not have headers, AT added a header with the Note’s title inside a <head> tag.
- 4) Although nested <bibliography> tags are permitted in EAD, the nested tags were imported as separate Bibliography notes.³⁰

E. *Access Control*

The final major data change during the the import was within <controlaccess>. The original file contained three <controlaccess> fields, which were structured as paragraphs with subject, person, and corporate names. Archivist’s Toolkit imported all the data within <subject>, <persname>, and <corpname> into the “Names and Subjects” section with @source=ingest. The data in the paragraphs was discarded.

While it is important to note all these issues and to strongly encourage archivists considering doing an import to attempt a few first and discover any major differences, the majority of the file imported without any loss of or change to the data.

³⁰ The effect achieved by the sample import’s <bibliography><p></p><bibliography>list of items</bibliography></bibliography> structure, in which <p> contained information about the bibliography’s contents could have been achieved in AT using the Note field inside the Bibliography note entry box. The structure would have exported as <bibliography><p></p>list of items</bibliography>. But while it is possible to achieve this in AT, that would still require the importing archivist to do additional work after import.