Hi, I'm Jody DeRidder, and I'd like to tell you about a recent NHPRC-funded project in which we developed a cheap, fast model for getting large manuscript collections on line.

Like other delivery methods that leverage the EAD for access to digitized content, we depend upon the information in the finding aid to for search and discovery. The better the finding aid, the better the access. While our grant-funded project included descriptions down to the folder level, only the series level descriptions are absolutely necessary. We have all this great metadata, created by the archivists, providing context for delivery. Why not use it? Most of us do not have the resources to get large manuscript collections online any other way.

Here's the information we had at hand, which could be transformed into item-level records, automatically. For others who want to use our system, we developed configuration files where you would enter the information on the left side for each collection; the information on the right side comes from the file names.   This is actually quite a bit of metadata. By creating minimal MODS records, we enable access to discrete items, and pave the way for later remediation, possibly by harnessing crowd-sourcing technologies.

Here's an example of how we encoded information in the file names, to avoid spreadsheets or hand-created metadata for each item.

- The filename starts with a letter so we can use the identifier as XML id attributes. The first segment of our file names indicates the repository source and type of content.
- The second segment is the collection number, echoing the one used by the archivists.
- The third segment has three parts: Box, Folder and sequence of the item.
- If there's a fourth segment to the identifier, that means there's more than one page, and this holds the page sequence.

In our archive and our web directories, the location of the file can be determined by replacing underscores with slashes; this helps us automate our work, and leverage the file system to organize our content. Our delivery system, Acumen, infers the relationships between the files by using these file name segments.

```
<titleInfo displayLabel="title">
        <title>TITLE_HERE</title>
</titleInfo>
<typeOfResource> %content_type% </typeOfResource>
<physicalDescription>
        <form authority="marcform">electronic</form>
        <internetMediaType> %media_type% </internetMediaType>
        <extent> EXTENT_HERE </extent>
        <digitalOrigin> %digital_origin% </digitalOrigin>
</physicalDescription>
<identifier type="local">FILEID_HERE</identifier>
<identifier type="uri">ITEM_LINK</identifier>
<location>
        <physicalLocation type="collection" displayLabel="Repository">
%repository_name% </physicalLocation>
        <physicalLocation type="collection" displayLabel="Repository Collection">
%physical_collection% </physicalLocation>
        <physicalLocation type="collectionNumber" displayLabel="Collection Number">
%collection_number% </physicalLocation>
        <physicalLocation type="boxNumber" displayLabel="Box
Number">BOX_HERE</physicalLocation>
        <physicalLocation type="folderNumber" displayLabel="Folder
Number">FOLDER_HERE</physicalLocation>
                        [etcetera]
```

MODS Template

This is a segment of the template we provide with our open-source software. The values with percent signs are filled in with the information provided in the configuration file for the collection. The capitalized segments are filled in from the file names… though the item link is actually a combination of the two. To use our software, you make a collection-specific copy of the basic config file, add in information about your specific collection and display requirements, add the path to your files, and run the scripts. The software will generate the MODS, link the content into the EAD, create HTML display files to meet your specifications, generate derivatives, and move content into your web directories.
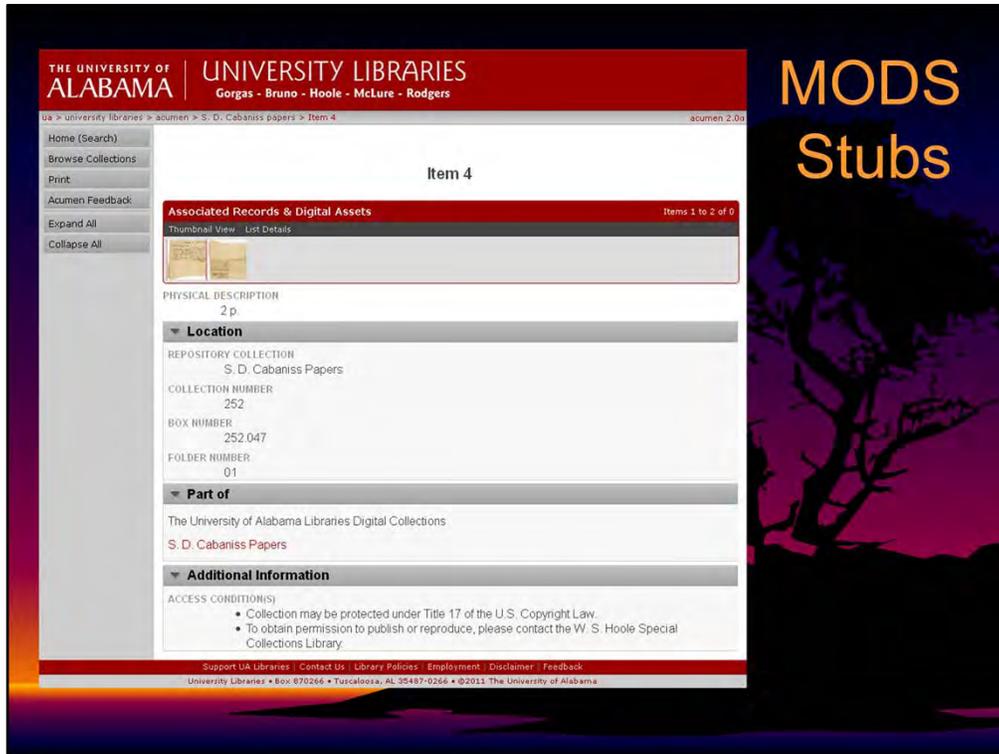
Here's an example of our own MODS display in Acumen. Clicking on the thumbnails provide large image access with zoom and pan capabilities.

However, we recognize that many people already have their own method for delivering EADs to the web. So for those who don't want to use our open-source Acumen software, we developed HTML templates for both items and folders which make use of your configuration specifications for logos, color, display and more.
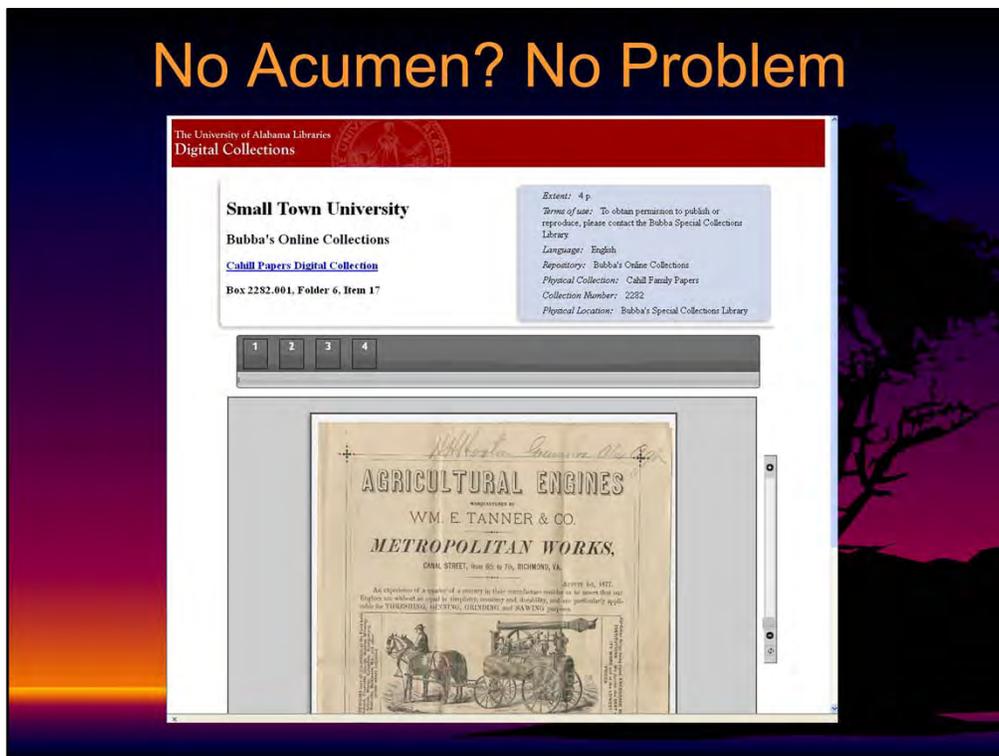
Here's a sample of an item using our HTML template. As you can see, the metadata appears in the upper portion of the page, and you can see at a glance how many pages the item has, and can access them via the navigation bar at the top.   This display also has pan and zoom capabilities. Of course the logo and information are all provided by you, so the actual display would be specific to your institution.
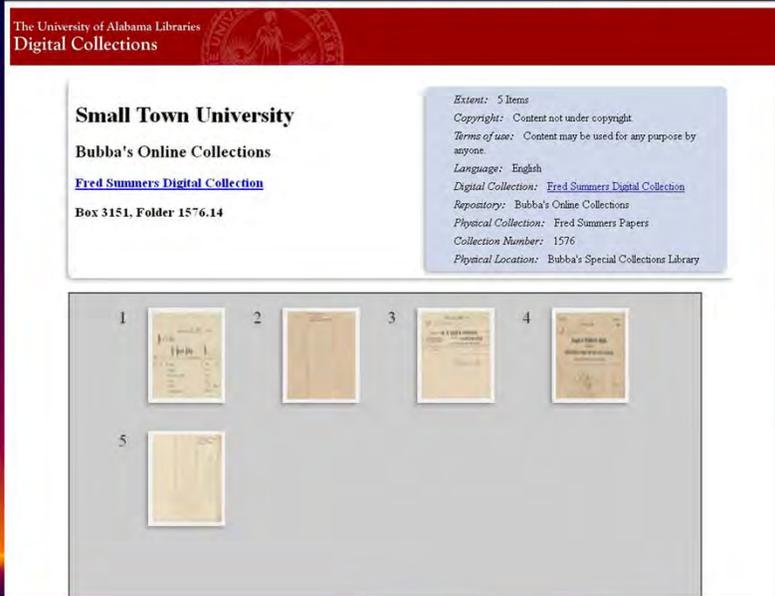
If you don't want to link a list of items into the EAD, you have the option of linking folders instead. This is preferable particularly if you do not have folder-level description.
Here is a sample display of the contents of a folder that was linked into the EAD. Clicking on any of the items brings you to the item page we just saw.
Again, known metadata is displayed at the top, with a return link to the EAD online to provide context.

Our focus was on streamlining and automation. The more we can automate our work, the cheaper it is.
We use scripts for almost everything, from checking the file names, verifying the EAD is linkable, generating the JPEGs, linking the content, and moving the files.

 Oh, and work can be completed in batches. For large manuscript collections, you want to make content available online as soon as you have it digitized, right? We support batch processing of segments of the collection at a time.   Digitize a box or two, get it online, and move on to the next boxes. No problem.

We recently analyzed the cost of delivering content this way; not counting hardware, software, supervision and overhead, it came to less than 80 cents per page. That's less than a third of our usual method of digitization, where we hand-create item level metadata. Clearly, this is potentially a great solution for large manuscript collections that may otherwise never see the light of day online.

We thank the NHPRC for funding this project.

Here's a link to a recent article about our work processes, links to our wiki, project site, and display; and also a link to our open source delivery system, Acumen, which automatically indexes content in web directories anywhere, and handles multiple levels of granularity, materials, and XML metadata simultaneously.

Please feel free to contact me with any questions.
Thank you!