I am Chris Lacinak from AVPreserve, and I'll be talking about the American Archive Archival Management System (AMS)
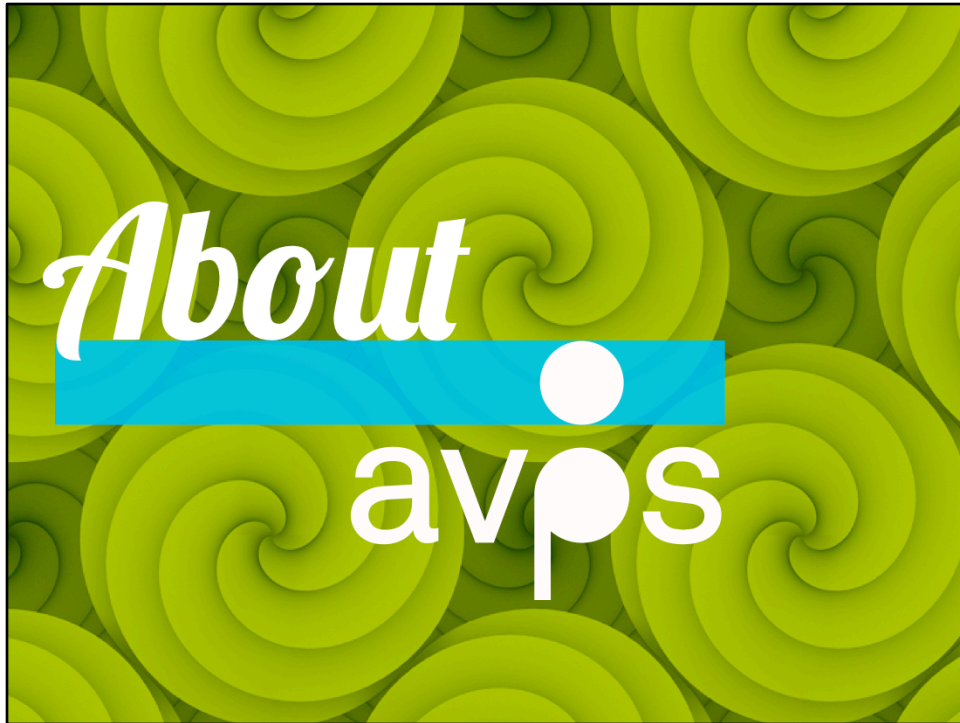
AVPreserve is a consulting and software development firm with three primary areas of focus. On one end of the spectrum we work on collection assessments and inventories of physical audio, video and film collections akin to the effort that Casey just talked about. On a different part of the spectrum we consult on digital asset management, digital preservation and metadata. And our third area, which has me here today, is software development where we address needs across the entire spectrum of consulting work we do, targeting custom development where needs are not being met by existing solutions.

AMS Background

At its core, the AMS is essentially the inventory data set with an application built around it to serve 3 main users:

- American Archive Admin
- Digitization Service Provider (Crawford)
- Stations

The AMS is the heart of the operation, in that it is where everything and everyone comes together. At a high level the AMS serves several functions, including:

Search, retrieval and playback
Importing and transforming metadata - from stations and service provider
Registering items (generating GUIDS)
Adding and editing metadata
Tracking and reporting on project status and key metrics
Exporting data
Communication with stations – messaging

It is important to recognize that the AMS is a metadata management system. It does not manage the media itself. It links to the media and renders it but it is not managing the media.
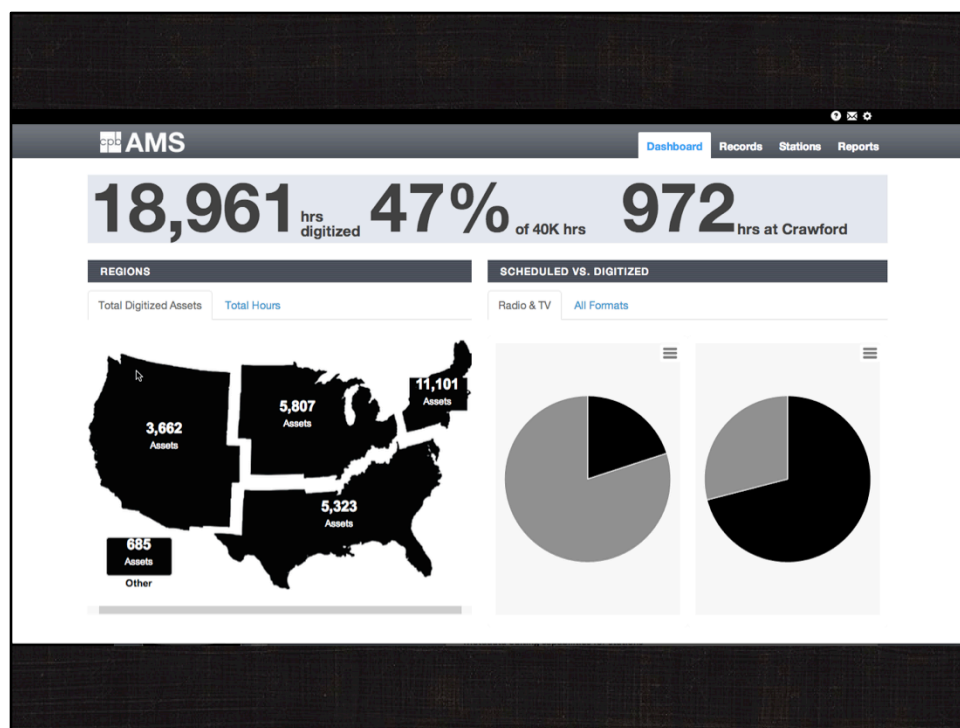
This shows the Administrative Dashboard which highlights
- the number of hours and items digitized
- The percentage of completion
- The number of hours that have been shipped to Crawford but not yet digitized in order to help maintain adequate backlog
- The number of hours/items digitized by region
- The percentage of scheduled vs. digitized items broken down by Radio and TV

Across the top you see we have 4 main pages consisting of Dashboard, Records, Stations and Reports
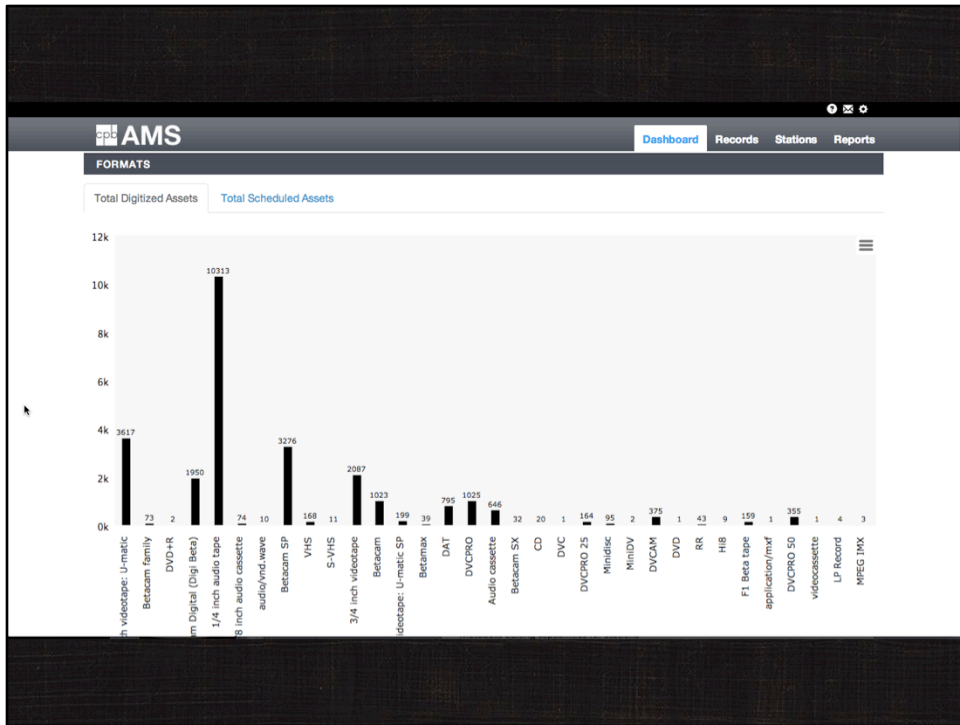
Scrolling down the Dashboard page we have reporting on a reporting by format of scheduled and digitized items
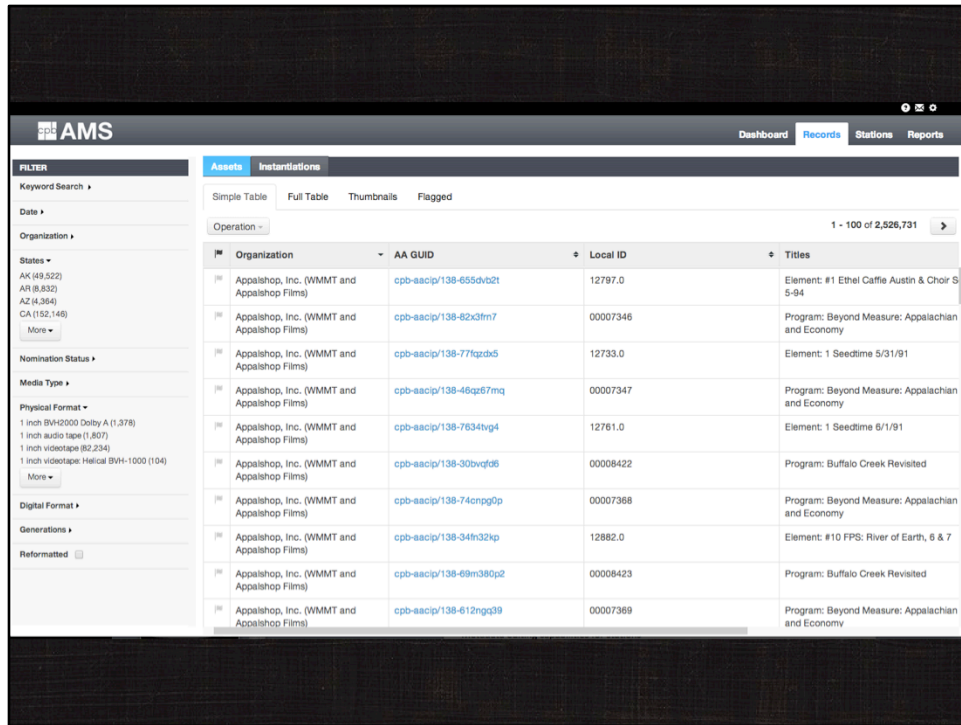
This is an Admin's main landing page for the Records. There are a few things to note here:

1. Highlighted in blue is text that says "Assets" to it's right is a tab called "Instantiations". We'll see more about that in a moment, but think of Assets as a given piece of content and instantiations as physical manifestations of that content.

2. On the left hand side we have keyword search and faceted browsing. Faceted fields include:
Date – of all date types in the system
Organization
State
Nomination Status
Generations and more.

An important note about nomination status. You'll see at the top right of the table a number in bold. That's the total number of records in the system. Those are all of the items that were documented as part of the inventory project. The first wave of digitization was restricted to 40k hours, and so each station got an allotment of hours for digitization. The way they identified their priority items for digitization was through nominating items.
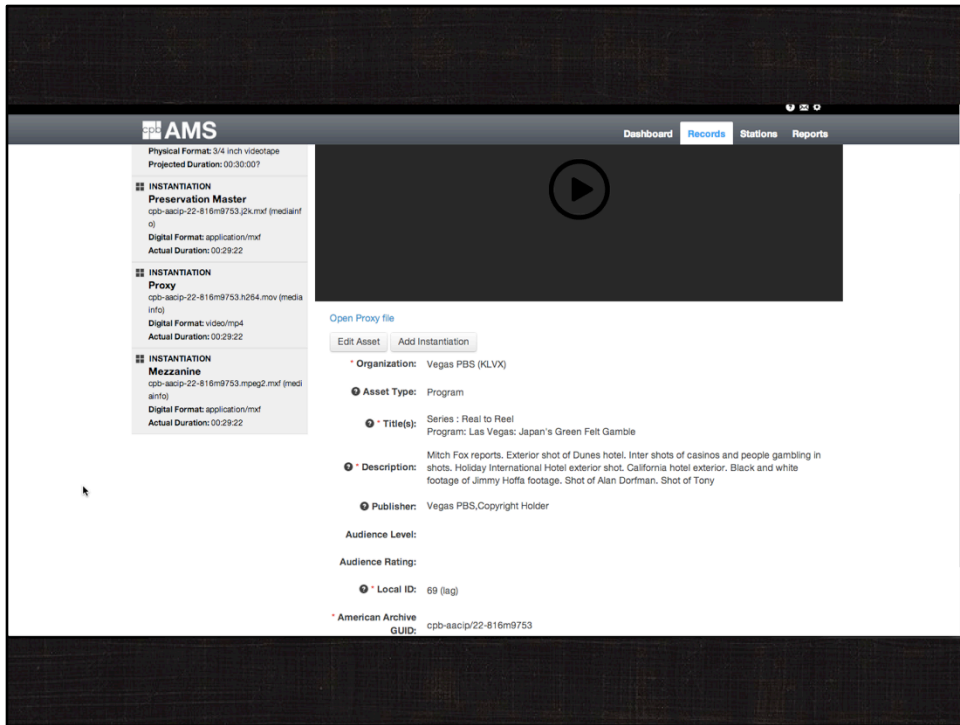
You'll see in the top left the filter criteria that has been applied. This is showing all reformatted/digitized moving image items from the state of Nevada so far. There are 70 of them

And if we click on one of them we'll go to this detail page.

What is highlighted in blue at the top there is "Asset Information". This is currently selected, and this dictates the metadata that shows up beneath the video. This is descriptive and administrative metadata that applies to all Instantiations.

Scrolling down we see fields ranging from the Local identifier to the Audience Rating.

We have now selected the Master which is the original ¾" Umatic tape and the metadata we see beneath the video now is mostly technical metadata describing this particular instantiation. The video remains the same throughout of course. It is just the same proxy to be able to listen and view the content.
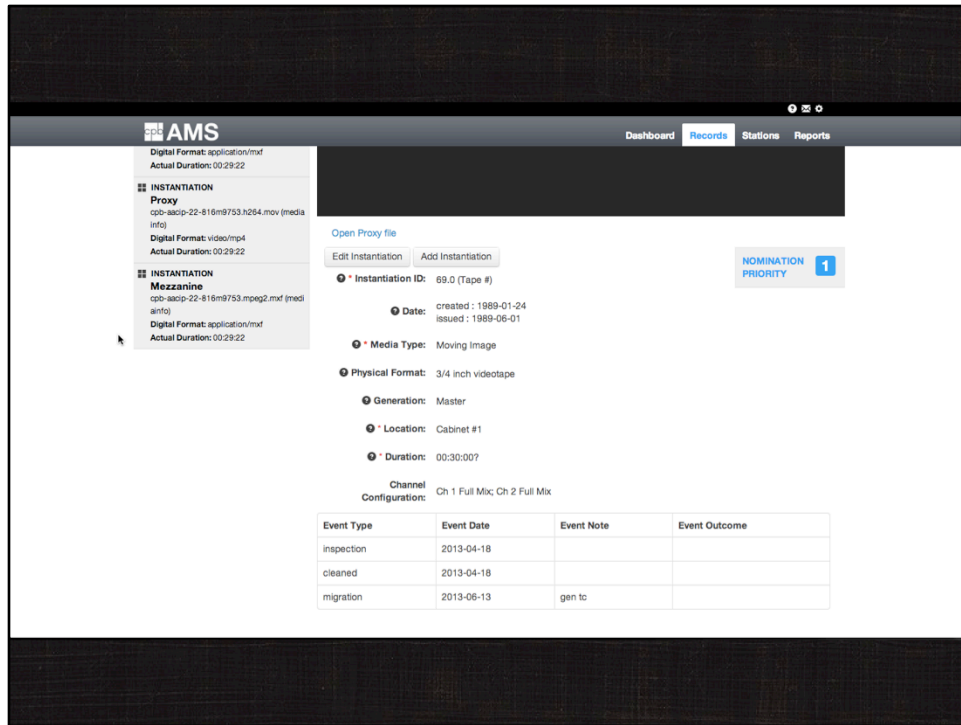
Scrolling down we see a few things

One is the nomination priority 1 over to the right. I talked about this earlier.
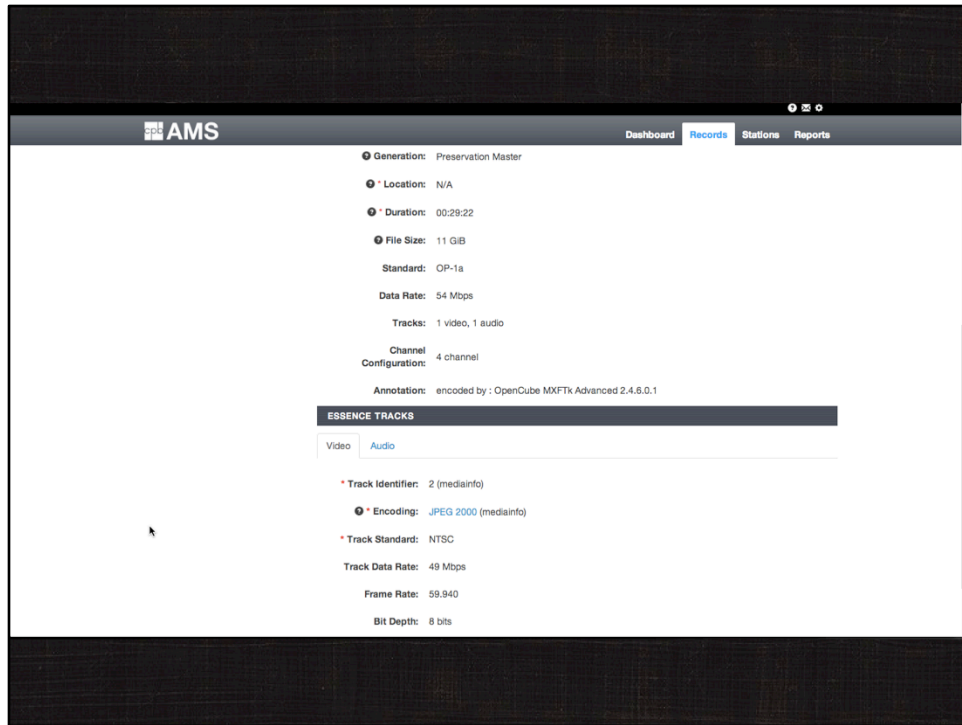
The other is the ability to add and edit instantiations.

The other is the table at the bottom. This is what is referred to as provenance or process history metadata. It tells us about the salient events that have taken place to a given instantiation. The metadata that is in this table is automatically extracted from Google Docs at Crawford using the Google Docs API and imported into the AMS. This is considered important preservation metadata.

This is the preservation master instantiation

Showing the result of importing the MediaInfo file generated by Crawford as part of their process in order to manage detailed technical information about a given file

And we also see a mezzanine and proxy file instantiations here too.
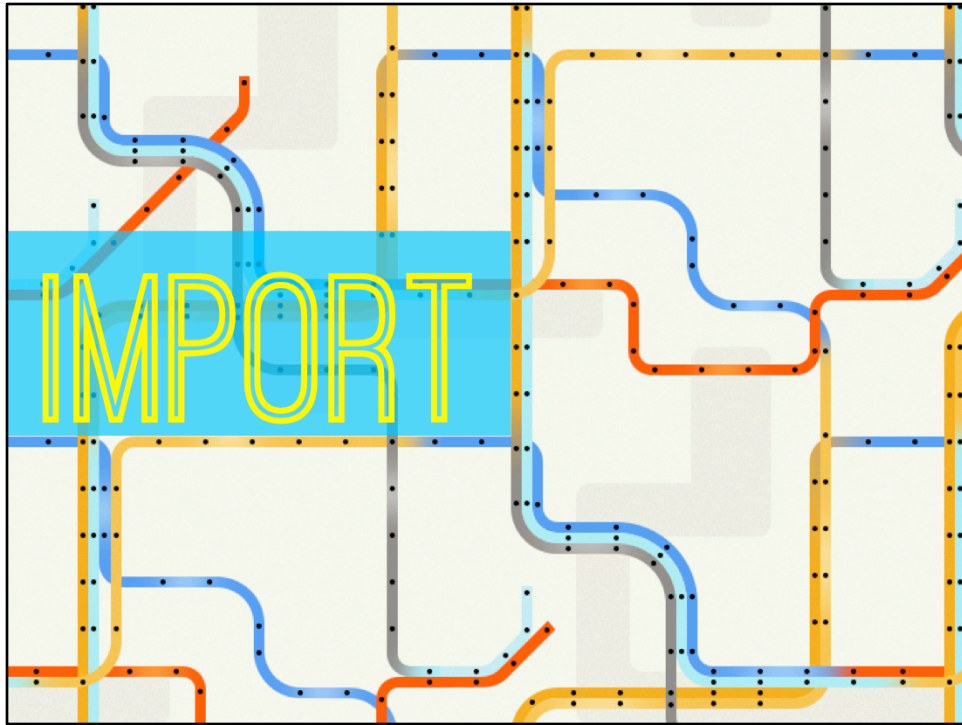
A really critical component of this project was enabling organizations to both import and refine their metadata. On the importing side

From within the records page you can go to the Operation menu and select "Import Collection"
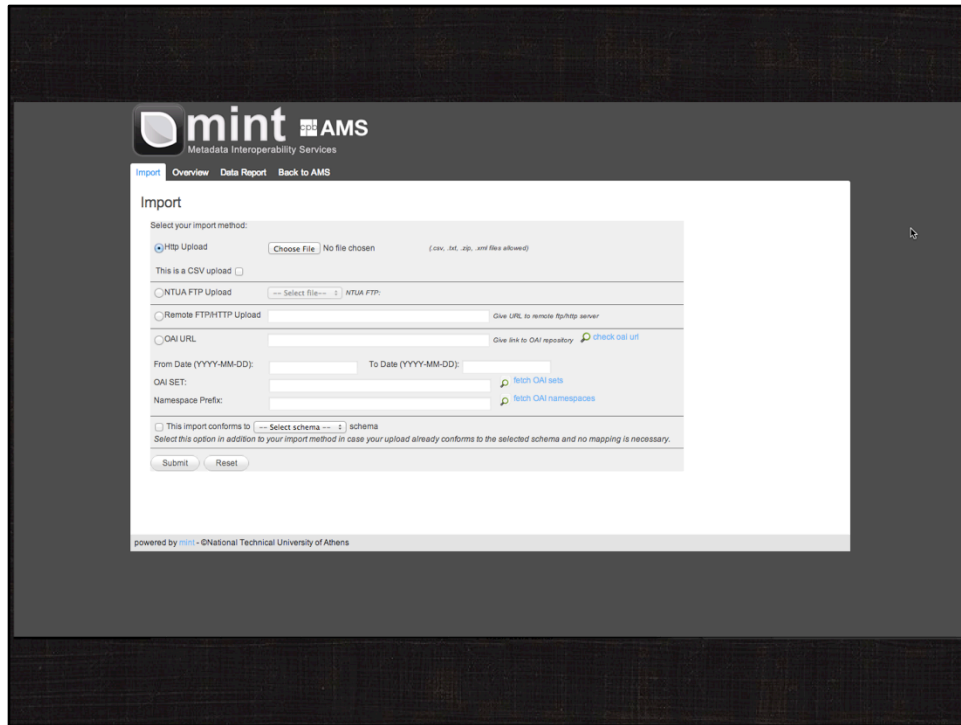
This will bring you to this interface which is Mint. Mint is an open source project out of Europe that is one of the more sophisticated tools out there for importing and transforming data. You can import CSV or XML
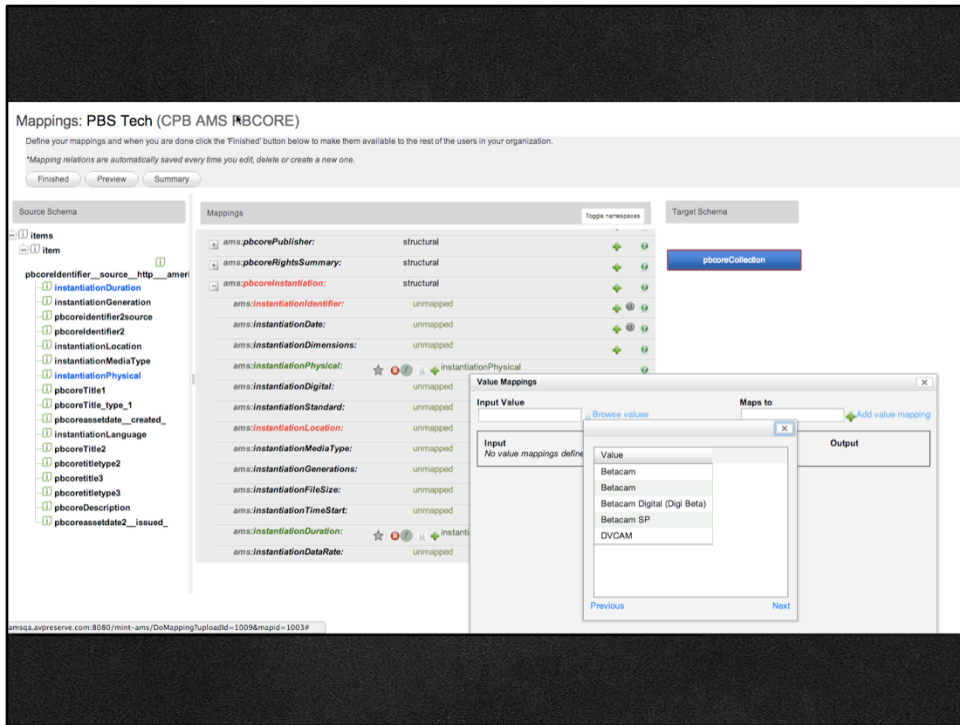
And this shows the mapping interface. On the left we see the the data that was imported and on the right we see the XSD. This window in front is showing that there are several functions you can do such as conditional actions, concatenation, normalization and more.
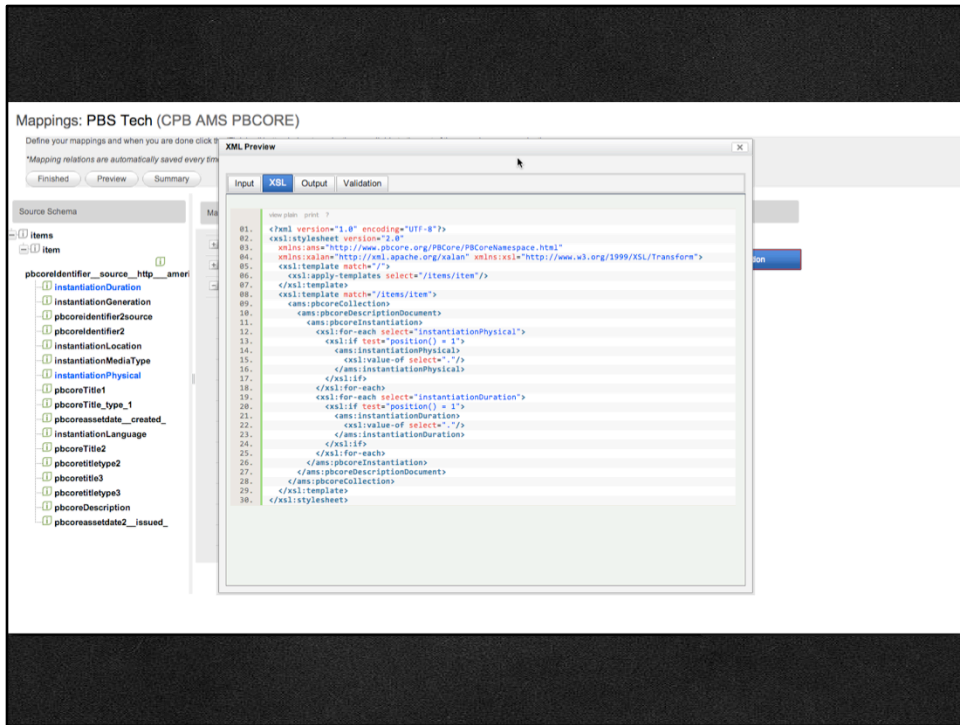
This shows the result as XSL. This same window will show you the input data, the XSL, the output or result and validation.

Moving on from import and transformation

Here we see a filtered set of records. If you look in the top left you will see that this is all results for Nevada, Moving image assets. In the top right of the table we see that this filtering results in about 6.5k assets. In the case of refinement, this filtering serves the purpose of selection. So filtering down and then selecting Refine data will only impact the filtered set.
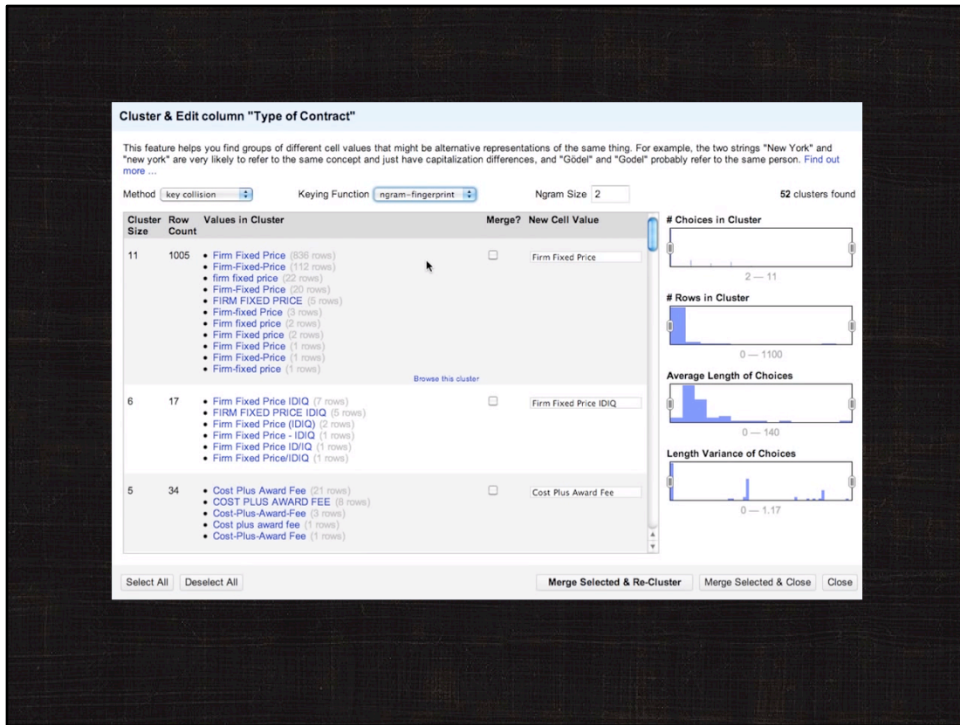
For refinement we integrated Open Refine, an open source data refinement tool from Google. This is not the American Archive data set in this image, but here we see an example of normalization. Refine has identified several cells containing similar data and provides the user with the ability to easily transform them all to a single expression.
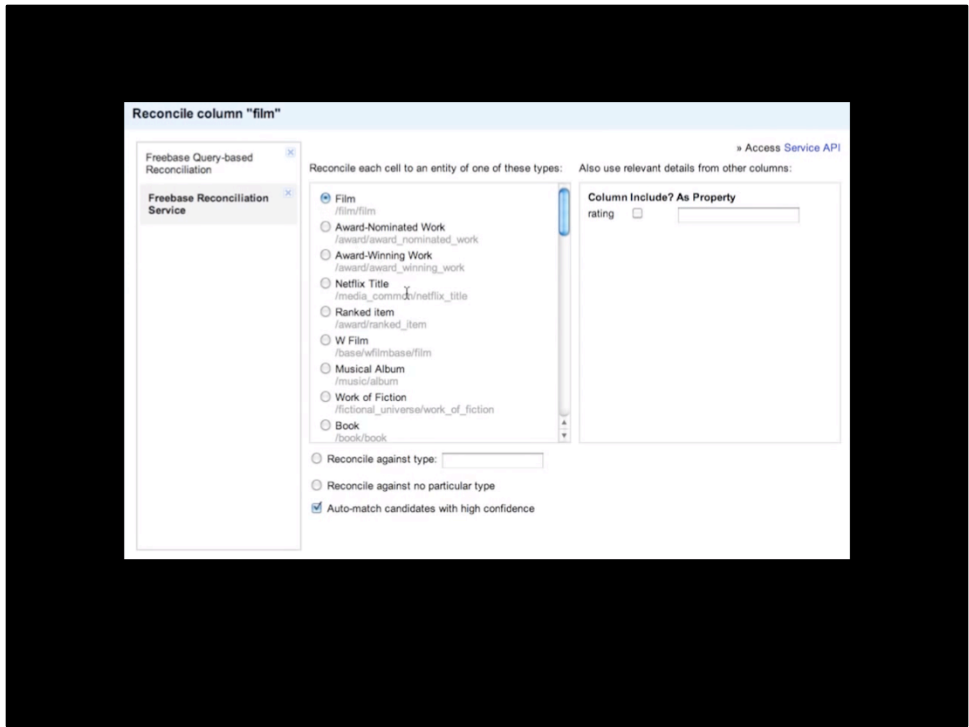
This is an example of Open Refine's ability to integrate with linked open data – in this case free base – based on named entities within the original data set.

Another big issue is exporting, and here the AMS offers two primary mechanisms for exporting

The first is from within the main interface. Again, filtering down to a set of results and selecting export results will export
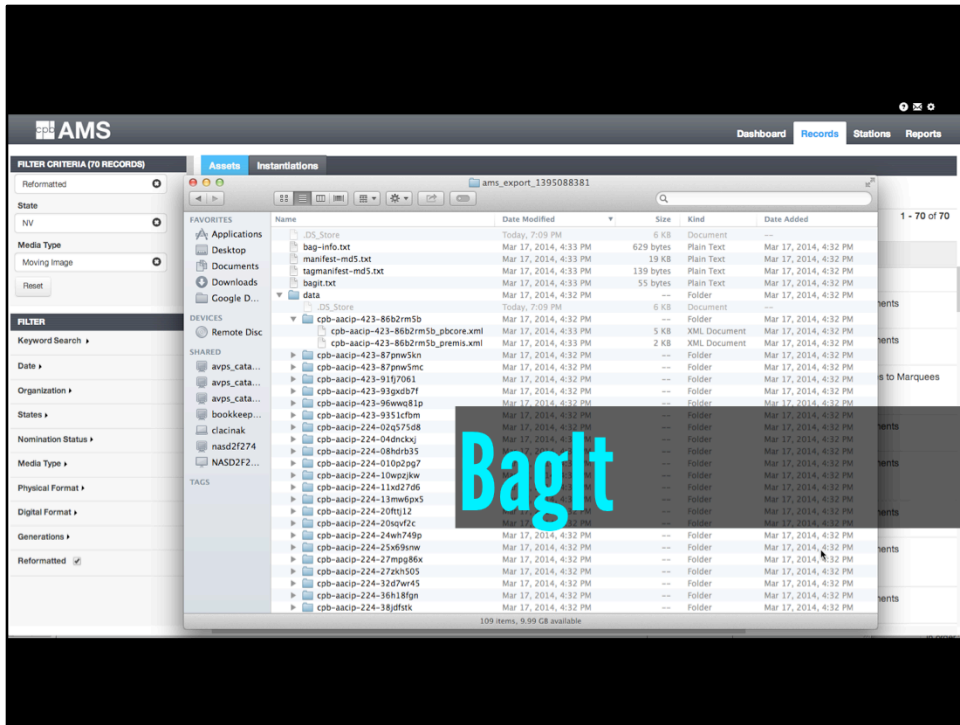
Those records as a bag, using the BagIt specification created by the Library of Congress and widely adopted as a packaging and exchange method. A bag is simple, and essentially contains a manifest of all the files it contains and checksums for each of the files. Each directory is named with it's GUID and within each directory is a PBCore XML and and a PREMIS XML document. For those of you that aren't familiar with PREMIS it is Preservation Metadata and primarily contains the event type data that we saw earlier in the item detail showing the events that have taken place to an item.

```
▼<pbcoreCollection>
  ▼<pbcoreDescriptionDocument xmlns:xsi="xsi" xmlns="http://www.pbcore.org/PBCore/PBCoreNamespace.html" xsi:xmlns="htt
    xsi:schemaLocation="http://www.pbcore.org/PBCore/PBCoreNamespace.html http://www.pbcore.org/xsd/pbcore-2.0.xsd">
    <pbcoreIdentifier source="http://americanarchiveinventory.org">cpb-aacip/83-49g4fmqx</pbcoreIdentifier>
    <pbcoreIdentifier source="KUED">1288</pbcoreIdentifier>
    <pbcoreTitle>Ancient of Days</pbcoreTitle>
  ▼<pbcoreDescription descriptionType="Abstract">
      3 episodes of "Alexander Schreiner at the Tabernacle Organ"
    </pbcoreDescription>
    <pbcoreGenre>Music</pbcoreGenre>
  ▼<pbcoreCreator>
    <creator>Keith Engar</creator>
    <creatorRole>Producer</creatorRole>
    </pbcoreCreator>
  ▼<pbcoreContributor>
    <contributor>Openshaw, Byron</contributor>
    <contributorRole>Director</contributorRole>
    </pbcoreContributor>
  ▼<pbcorePublisher>
    <publisher>KUED</publisher>
    <publisherRole>Copyright Holder</publisherRole>
    </pbcorePublisher>
  ▼<pbcoreInstantiation>
    <instantiationIdentifier source="KUED">1288</instantiationIdentifier>
    <instantiationPhysical>DVCPRO 25</instantiationPhysical>
    <instantiationLocation>KUED Production Archive> Row 11> Unit 1> Shelf 5</instantiationLocation>
    <instantiationMediaType>Moving Image</instantiationMediaType>
    <instantiationGenerations>Dub</instantiationGenerations>
    <instantiationDuration>01:27:48:00</instantiationDuration>
    <instantiationChannelConfiguration>Ch 1 Full Mix; Ch 2 Full Mix</instantiationChannelConfiguration>
    <instantiationAlternativeModes>N</instantiationAlternativeModes>
    ▼<instantiationExtension>
      ▼<extensionWrap>
        <extensionElement>AACIP Record Nomination Status</extensionElement>
        <extensionValue>Nominated/1st Priority</extensionValue>
        <extensionAuthorityUsed>AACIP</extensionAuthorityUsed>
        </extensionWrap>
      </instantiationExtension>
    </pbcoreInstantiation>
  ▼<pbcoreInstantiation>
    <instantiationIdentifier source="mediainfo">cpb-aacip-83-49g4fmqx.h264.mov</instantiationIdentifier>
    <instantiationDate dateType="encoded">2013-07-30</instantiationDate>
    <instantiationDigital>video/mp4</instantiationDigital>
    <instantiationStandard>QuickTime</instantiationStandard>
```

The other method of exporting data is using web-services that were built based on GUID, Digitization date, and last modified date in order to retrieve both PBCore and PREMIS XML output. This is a better method for using automated mechanisms of export whereas the previous method of export is oriented toward manual export.

And that's all I have time for today. Thanks.