

Spreadsheet Madness: Mapping Data from the USC Shoah Foundation Visual History Archive to MARC

Daryn Eller

Description: The 50,000-plus audiovisual testimonies in the USC Shoah Foundation Visual History Archive each have significant metadata, but it is organized in ways unique to the Institute and not compatible with Library of Congress standards. When tasked with creating library records for the archive, the challenge was to not only map the individualized data behind each testimony to MARC, but - since viewing each testimony for content would be impossible - to come up with uniform subject headings and summaries that could be used across experience groups. The presentation will take conference attendees through the process, familiarize them with the tools used, and explain how it was attempted to make the records less generic by adding geographical terms specific to each testimony. Also covered will be the additional challenge of getting the MARC records into USC's Primo/Alma catalog and the collaboration with USC librarians. Is it a good idea to create such generic records? Will it help researchers find useful resources or clutter up their searches?

Daryn Eller is assistant archivist at the USC Shoah Foundation, where she manages the institutional media archive. She has a BA from University of California, Berkeley, and an MLIS from San Jose State University.

PowerPoint Script

FIRST SLIDE: Hi Everyone and thank you for coming to hear about Spreadsheet Madness at the USC Shoah Foundation. It's a tale of mapping proprietary data to MARC records and all the steps and challenges in between.

Sandra has just introduced you to the Visual History Archive, which we call the VHA. I'd like to tell you about a project we began in 2018 with the goal of increasing awareness of the Shoah Foundation's collection of testimonies by making them discoverable in library catalogs.

SECOND SLIDE: Most people know who Steven Spielberg is and many of them even know that he started a foundation dedicated to collecting the testimonies of Holocaust survivors.

THIRD SLIDE: But there are still many people who don't know anything at all about the USC Shoah Foundation's testimonies or how to access them. There are also many people who don't know how the Shoah Foundation's collection has expanded. They don't know, for instance, that among the many different types of oral histories we have in our archive there are over 300 testimonies from the Armenian genocide and over 100 testimonies from the Nanjing massacre.

FOURTH SLIDE: This isn't to say that the Shoah Foundation's archive is obscure. A recent assessment found that, in a one-year period, over 173,000 people searched the collection at one of our access sites

or through our online portal. Nonetheless, the content of the archive is extraordinary and, of course, we want the VHA to be an even more widely used resource.

FIFTH SLIDE: So how could we get the word out?

SIXTH SLIDE: As librarians, you know what the key to greater discovery is: catalog records. Sam Gustman, the institute's Chief Technology Officer--who is also associate dean at the USC Libraries--tasked our team of archivists and a curator with creating MARC records for each of the testimonies in the VHA. The team was led by Sandra Aguilar, who you just heard from, and the implementation was done by myself and Svetlana Ushakova.

SEVENTH SLIDE: Here you see the interfaces of our two initial target catalogs: the USC Libraries catalog and OCLC's WorldCat. Our plan was to first make the records available to the public through these two catalogs and to provide links in the records that would drive users to the Visual History Archive.

EIGHTH SLIDE: If the users were at one of our 147 access sites, they could watch any testimony with one click. What you're seeing here is the link in a USC record and where clicking on the link sends the user. If a library patron is looking at the records from a place that is not an access site, that person would still be able to see biographical data about a testimony and find out where to view the video. There are 3,000 of the testimonies available online through any web browser, so the user could potentially even view a video without being at an access site.

NINTH SLIDE: As we proceeded to put this plan into action, we faced two main challenges. The first was that, although we had plenty of descriptive metadata about each testimony, it didn't fit neatly into the prescribed MARC fields. We would, in essence, have to make a square peg fit into a round hole.

TENTH SLIDE: The second problem was that we had to create approximately 52,000 records. These records, of course, would have to be created in bulk, which meant, unfortunately, that we couldn't describe each testimony in detail or give subject headings specific to each videorecording. We were concerned that the records would be too generic—the library version of a Brand X cereal.

ELEVENTH SLIDE: I'd like to first describe how we handled challenge number 2. The scope of the project involved making records for 18 different experience groups. The archive, for instance, not only has testimonies from Jewish survivors, it has testimonies from survivors who were Jehovah's Witnesses and people who participated in War Crime Trials. Although we had to come up with standardized summaries and subject headings for each of the experience groups, we tried to make them as descriptive as possible and to differentiate the experiences. This slide compares the summary and subject heading fields for a Jewish survivor and a Sinti-Roma survivor and you can see that there are some differences. The descriptive information is still very basic, but it will still help users narrow their searches. And because the records link to the Visual History Archive, once users get there, they will be able to use the VHA search tool to do very detailed searches.

TWELTH SLIDE: To enrich the records further, we decided to add geographical subject headings to some experience groups. These geographical terms are related to places that survivors spent a lot of time in

and/or talk about extensively in their testimony. Putting these terms in the records will help researchers and other users discover testimony from people who were, say, imprisoned in a certain concentration camp or ghetto, or who come from a specific town. What made this process challenging is that the geographical terms in the Visual History Archive don't always conform to Library of Congress name authorities. This slide shows how the town of Armiansk is listed in the Shoah Foundation thesaurus versus how it's depicted in the Library of Congress name authority. We have had to map our terms to Library of Congress authorities and it's been a very time-consuming effort. Nonetheless, we hope to eventually add geographical terms to all the records.

THIRTEENTH SLIDE: The answer to the first challenge I mentioned—how to fit the square Shoah Foundation data into the MARC round hole—was solved through the use of spreadsheets. That's when the madness began. We started with an Excel spreadsheet of Shoah Foundation testimony data, then we mapped the data to another spreadsheet, this one with MARC fields. The process involved using a number of different formulas to slot everything into place, and to add in terms that would make the data readable.

FOURTEENTH SLIDE: Once our spreadsheets were in order, we turned to the MarcEdit tool to help us create xmls for loading into the USC Libraries Alma/ExLibris catalog system. This slide illustrates the many steps it took to take the data from a spreadsheet to xmls in MarcEdit. The tool worked very well and once we got the hang of using it, the process went fast. MarcEdit also allowed us to validate our records.

FIFTEENTH SLIDE: Once the records were uploaded into the USC Libraries catalog we had some kinks to work out, which we did with the help of the library staff in charge of the catalog. Most of the problems had to do with display and required some adjustments to the Primo/Alma system used by USC libraries. When the fixes were in and the records were displaying well, the library shared them with OCLC's WorldCat. We anticipate that the number of institutions incorporating our Shoah Foundation MARC records into their catalogs will grow with time. In the upcoming months, the records will be made available to Alma subscribers through a managed community zone. We know that Georgetown University has already downloaded the records and that at least one international university plans to move them into its catalog as well. We have also shared the records with Cornell University.

SIXTEENTH SLIDE: Library catalogers, of course, can never rest on their laurels. There is always room for improvement and records must reflect changes in collection. With that in mind, we made a four-point plan for going forward. Our first action was Follow-up – We made a point of checking to see if the records were displaying properly and the links were linking, not just in the USC Libraries catalog, but in WorldCat and the Georgetown Libraries catalog. In some cases adjustments had to be made. Our second action will be to Update – We created a schedule to refresh the records every six months to reflect changes that periodically occur in the Visual History Archive. The third task is to refine the records – As we refresh them, we will also refine the records by adding more geographical subject headings. The last factor in our follow up is Feedback.

SEVENTEENTH SLIDE: We want to hear from librarians like you on how the records read, link and display. If you're so inclined, we hope that you'll take a look at them on the USC libraries website. Thank you so much for listening to our spreadsheet-MARC record story.

EIGHTEENTH SLIDE: Any questions?

NINETEENTH SLIDE: Our contact information.