

Technical Assistance Consultancy Program - Final Report
Submitted by Richard Shrake
May 30, 2017

Overview of Project:

Vermont Folklife Center (VFC) Archivist, Andy Kolovos, worked with consultant Richard Shrake, a Vermont-based digital archivist and database/web developer, to achieve two goals:

1. Export data from legacy database, import data into VFC Omeka-based digital archive.
2. Update VFC's Omeka-based online digital archive to the most current version.
3. Evaluate options for, and select, a new archival cataloging system. Assist with transfer of legacy content into new system.

1. VFC Legacy Database

In 2002 the Vermont Folklife Center was awarded a grant from NEH to digitize parts of its collection and create an online access system to make these materials remotely available. The resulting custom MySQL database remained functional in varying degrees until 2017.

Goals

Export data from legacy MySQL database, combine relevant data from data tables into single CSV file. Structure fields in CSV file to correspond with Dublin Core. Import CSV into VFC Omeka.

Work Undertaken

Evaluation of data in legacy system

Shrake proceeded to evaluate the data held in disparate MySQL tables from a long-dormant, in-house management system. After deciphering the structure of these tables, Shrake exported multiple XML files and developed a series of XSLT stylesheets to bring the data together for a single csv representation of a collection for uploading into Omeka. With slight alterations, these forms will permit further export of data from an otherwise inaccessible system as needed.

Import of legacy data into Omeka

Using Omeka's CSV Import Plugin, Shrake imported data into VFC Omeka system.

Remaining Work

Kolovos will review all imported records for completeness and correctness and link associated audio, image and text files to them.

2. Omeka Introduction:

VFC maintains an online digital archive (<http://explore.vermontfolklifecenter.org/digital-archive/collections/>) built using the Omeka platform. Omeka is a free, open source database system developed by Roy Rosenzweig Center for History and New Media, George Mason University and employed in museums, archives and libraries to present digital collections online. VFC's Omeka instance is hosted via a free, non-profit shared-hosting account at the Dreamhost ISP.

Goals

Upgrade Omeka version 2.0.1 to Omeka version 2.5. Ensure integrity of data in the transfer. Ensure functionality of the system post upgrade. Review currently installed plugins and add new plugins as needed/desired.

Work Undertaken

System Upgrade

Creating collections in Omeka is a fairly straightforward and intuitive process, but working with the structure of the software requires the skills of a web and database developer. After several years using an early version of Omeka, VFC found it necessary to upgrade their installation to the most current version of the software, Omeka 2.5.

The process of updating Omeka is well-documented, but can still cause consternation since it involves migrating, deleting, and updating directories and updating a MySQL database. Shrake initially attempted to perform a software update directly from Omeka 2.0.1 to 2.5, but encountered persistent SQL errors. Performing an intermediate update to 2.4.1 and then a final update to 2.5 proved successful.

Kolovos and Shrake tested the system post-upgrade and determined that the upgrade had been successfully executed.

Plugins

Omeka users have created a wide array of plugins that enhance and expand the functionality of the system. After Shrake completed his upgrade, Kolovos undertook the process of reviewing installed plugins to ensure compatibility with the newly updated version of Omeka and identified and installed new plugins.

Remaining Work

In an effort to batch-correct data in the system VFC needs to export CSV files of relevant collections, revise these data using Excel, export CSV from Excel, and re-import the revised data into Omeka. Kolovos originally hoped the necessary batch changes could be executed directly in Omeka via a plug-in, but investigations by Kolovos and Shrake revealed the necessity of exporting, correcting and re-importing data. Using the process devised by Shrake in Item 1 (see above), Kolovos anticipates piloting this method in early June and, if successful, completing the work by mid-July.

3. Archival Cataloging System Introduction:

In approximately 2008 VFC adopted Archivist's Toolkit (AT) software to catalog their collections. AT was a free, open source, standards-based archival descriptive system developed by the University of California San Diego Libraries, the New York University Libraries and the Five Colleges, Inc. Libraries. In August of 2013 with the release of ArchiveSpace, all upgrades and support for AT ended. Despite the product's end-of-life, VFC continued using AT to describe materials and generate EAD-encoded finding aids. However in the last year VFC's AT database became inaccessible. With this problem serving as a catalyst, Kolovos started to look at other options for archival descriptive software.

Goals

Review options for archival descriptive software, select a system, transfer data from AT to the new system.

Work Undertaken

Shrake and Kolovos met to discuss VFC needs in a descriptive system and explored two options: Archon and ArchiveSpace. After reviewing the pros and cons of these two options, Kolovos opted to shift to ArchiveSpace. Shrake explored data migration options.

Selection Process

Shrake and Kolovos limited their review to Archon and ArchiveSpace ignoring commercial products that provided some similar functionality, such as Past Perfect, as well as more complex, integrated systems such as Islandora. Primary concerns for Kolovos were *Cost*, *Compatibility with AT*, and *Support*. Ultimately, as outlined below, VFC adopted ArchiveSpace.

Archon

Archon, like AT, ceased development and support in August of 2013 in lieu of the release of ArchiveSpace. However an active group of Archon users have continued to work on the

software on their own and make it available. Archon continues to be free and Shrake and Kolovos felt it could likely be hosted on the extant VFC Dreamhost account. Archon scored high in the areas of *Cost* (Free) and *Compatibility*, but in the opinions of Shrake and Kolovos fell woefully short in the area of *Support*. Since VFC lacked the skills in house, necessary ongoing maintenance would potentially be costly and user support would be limited.

ArchiveSpace

As noted above, ArchiveSpace was developed in part to replace both AT and Archon. While ArchiveSpace software is free, access to user support is limited to those who are paying members of the ArchiveSpace consortium. Annual membership rates, which are based on the total number of FTEs at an institution, can be quite expensive.

While anyone, member or not, can download, install and make use of ArchiveSpace at no cost, the expense of maintenance, upgrades and other technical support for self-hosted ArchiveSpace instances falls on the host repository. As with Archon, since VFC lacked the skills in house, necessary ongoing maintenance would potentially be costly and user support would be limited.

Annual membership in the ArchiveSpace Consortium for VFC would be relatively low, in the area of \$400, and although this would provide access to user support materials, VFC would still be responsible for costs related to installation, configuration and ongoing maintenance of any self-hosted ArchiveSpace instance.

Several services (e.g. LYRASIS, Atlas Systems) offer commercial hosting plans for ArchiveSpace that provide technical support and, in some cases, access to user materials as a part of the hosting fees. These hosting plans are generally directed at larger institutions and repositories and can cost on the order of several thousand dollars annually. Furthermore, pricing for these plans is not transparent--one must contact sales representatives at the various providers from whom one gets a quote for the annual cost of service.

These choices left Kolovos flustered and did seem to leave VFC stuck between a bit of a rock and a hard place between *Cost* on one side and *Support* on the other. Shrake and Kolovos stuck at it, and eventually Kolovos's scouring the hosting offerings at LYRASIS turned up a level of service they termed "Lone Arranger Hosting."

The Lone Arranger Plan of LYRASIS is a lower cost level of service that allows a repository a spot in a shared instance of ArchiveSpace. As a part of the service, LYRASIS provides access to user support materials and assumes responsibility for technical support. LYRASIS quoted VFC an annual cost of \$400.

The Loan Arranger plan has some fundamental limitations. Unlike the standard, more expensive, ArchiveSpace hosting plans which provide each repository with their own, unique

installation of the software, the Lone Arranger Plan provides space to multiple repositories in a shared installation of ArchiveSpace. As a result all the repositories using the same Lone Arranger installation share databases that hold many controlled terms, including names and subjects. While this co-mingling isn't deeply problematic, it just means that individual users must scroll through additional data to access their preferred, often-used terms. Kolovos and Shrake decided that the LYRASIS Lone Arranger Plan offered the best compromise between the identified VFC parameters of *Cost*, *Support* and *Compatibility*.

VFC then contracted with LYRASIS for the Lone Arranger Plan and began working with Shrake and LYRASIS to migrate legacy data from VFC's AT database.

Data Migration

As noted above, VFC was no longer able to access their AT database via the installed client on Kolovos's work computer. Kolovos, being of a pessimistic nature, simply assumed this was a critical failure, that the database was permanently inaccessible via the client, and requested that Shrake investigate and see what, if anything, could be done to export the data. Following some diagnostics and attempts to extract the data directly from the database, Shrake downloaded the final version of the AT client software and was able to access VFC's database without difficulty. Kolovos followed suit and replaced the old version of the AT client with the final version. He too was able to access the database.

Occam's Razor prevailed.

Kolovos then corresponded with LYRASIS support to determine the best path for migration, and Shrake began to assemble and test the standard AT to ArchiveSpace migration tools. That's when some additional limitations of the LYRASIS's Lone Arranger plan became apparent.

Due to the shared nature of LYRASIS Lone Arranger hosting the standard migration tools developed for AT could not be used. VFC would have to export EAD XML for each collection and import these files into ArchiveSpace individually. While this was not a deep complication, it also meant that all the names and subjects housed in our AT database would not automatically transfer to ArchiveSpace. Shrake will provide some assistance in developing macros to automate these ingests when VFC is prepared to migrate its data.