

case study

New York Philharmonic Digital Archives to Preserve and Make Accessible Worldwide Over 160 Years of American Cultural Heritage with the help of Technology Services Group and Alfresco Software

Customer

New York Philharmonic

Industry

Nonprofit

Solution

Document Management
Web Content Management

Products

Alfresco Enterprise 3.4.1
TSG OpenMigrate
Solr
Microsoft SQL Server 2008
Virtualized Environment
Windows Server 2008
GNU BookReader
ImageMagick

Partner

Technology Services Group
(TSG)

Results

- First three-year project phase includes digitization of 1.3 million pages of archival material from *The International Era, 1943 to 1970*, which includes 3,200 programs; 8,000 folders of business records; 4,200 glass lantern slides; 8,500 photographs; and 72 scrapbooks of fragile press clippings.
- Alfresco will hold 10 million nodes comprising 5 TB of data at the end of first phase.
- Implemented highly scalable content platform which will eventually hold entire digitized archives as well as born-digital records, audio and video (estimated 2 petabytes of data).
- SOLR implemented to create easy site navigation and instant, meaningful search results.
- Scalable content ingestion process using TSG's OpenMigrate toolset.
- Custom metadata structures to index content and provide institutional context of material.
- Clustered environment to ensure content availability 24 hours a day.



“ Plunging into the archive is a heady experience. It is full of unexpected gems and insight into musical life of a past era. ”

— Alex Ross, *New Yorker* critic
and author of *The Rest is Noise*

“ We have strong commitment to open source software and we have chosen platforms, such as Alfresco, that have strong developer communities and a wide user base. Our long term plan is to utilize Alfresco for digital asset management internally — everything from current concert recordings to email. Alfresco will be the single repository to bring together past, present and future content. ”

— Mitch Brodsky, Digital Archives Project Manager at New York Philharmonic

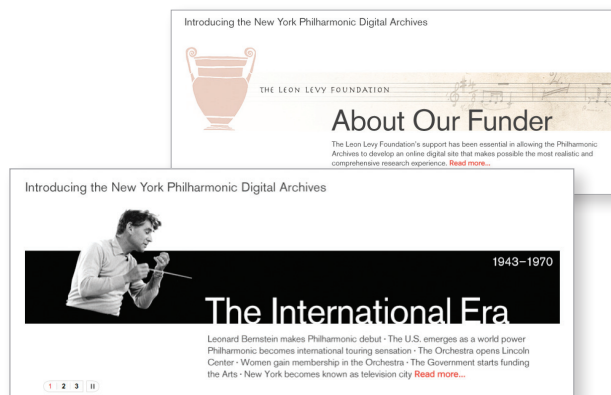
Challenge

Founded in 1842, the New York Philharmonic is the oldest symphony orchestra in the United States and the third oldest in the world. As such, the Philharmonic Archives is one of the oldest and most important orchestral research collections in the world. It traces the entire history of the Philharmonic and its more than 15,000 performances around the world and is an important record of cultural history in New York City and beyond.

In September 2009, the New York Philharmonic received a \$2.4 million grant from the Leon Levy Foundation to digitize 1.3 million pages of material from its archives, making them available to scholars, musicians, students, and the general public over the Internet. The Archives' collections contain material that dates back to the Philharmonic's first concert in 1842, but the first phase of the digitization project focuses on the Philharmonic's *International Era, 1943 to 1970*. This included digitizing 1,300 scores marked by Leonard Bernstein and Andre Kostelanetz, 3,200 programs, 8,000 folders of business records, 4,200 glass lantern slides, 8,500 photographs, and 72 scrapbooks of fragile press clippings.

The Philharmonic is one of the first institutional repositories to embark on a digitization project of this size and scope with the intent of making all digitized material available worldwide. In order to complete the project, it needed a highly scalable document management system that could handle heavy daily use while continuously streaming large volumes of data. The solution needed to be cost effective, handle large files and have strong digital asset management capabilities.

The organization focused on open source technology as it is easily scalable, reliable and cost effective. As well, open source offers more flexibility to create a solution that is sustainable over the long term and can be easily shared with other institutions.



Solution

The Philharmonic researched open source enterprise content management products evaluating Alfresco Enterprise and Fedora Commons, an open source digital repository framework commonly used for academic digital libraries. The team selected Alfresco because it offers a commercial product backed by support services, can easily scale for high volumes of content, supports any file type and has a robust developer community. In addition, Alfresco could serve as a content platform for the Philharmonic's born-digital archives and be customized to meet the organization's specific needs into the future.

To help implement Alfresco and streamline the content ingestion process, the Philharmonic turned to Alfresco Partner, Technology Services Group (TSG). TSG's OpenMigrate software controls the flow of all metadata and images into and out of the Alfresco repository,

allowing the Philharmonic to perform bulk metadata imports, image ingestion, and Web-enabling assets by indexing content in the front-end Solr search application. Content renditioning is performed prior to ingestion using a standalone implementation of ImageMagick, an open source software suite that converts the original JPEG images into web-optimized derivative files of various sizes.

The Philharmonic uses clustered Windows servers so that the image conversion and ingestion process can be scaled to meet even the most demanding schedule. Each day, some 120,000 jpeg images are ingested and up to 75,000 are deleted to make way for corrected replacement images. At the same time, the front-end site must maintain speedy content delivery for public use as well as internal content proofing. This level of demand requires a highly scalable system such as Alfresco in order to maintain accurate indexes while providing fast content retrieval and modification.

The asset viewer utilized for the final presentation of digitized assets is the open source GNU BookReader, started by the Internet Archive and now hosted on Open Library (openlibrary.org). The viewer allows users to pan, zoom, rotate, magnify, view thumbnails, and virtually turn pages. The painstaking detail in the Philharmonic's photography methods and quality control workflows allows end users to see more and do more with the digital asset than they ever could with the physical item on a table in the reading room.

“ Alfresco is the ideal content platform to handle the sheer scope and scale of the New York Philharmonic Digital Archives project. It provides a modern architecture with the flexibility customers need to manage massive amounts of data — whether it is for historical or business content. We are pleased to be part of this important undertaking to make valuable historic content available for everyone around the world to view and listen to. ”

— Dave Giordano, Founder and President at Technology Services Group, Inc.

Conclusion

- The New York Philharmonic Digital Archives will contain over 1.3 million pages of material from *The International Era, 1943 to 1970*.
- Alfresco serves as the content platform for the project and will hold 10 million nodes comprising 5 TB of data by close of the current phase.
- Digital Archives is freely accessible worldwide and provides comprehensive search interface with easy-to-use document viewer.
- In its first four months, the Digital Archives received 47,800 visits, 34,000 of those unique. 5,264 visitors returned to the site nine or more times, and 885 of those returned 100 or more times. Leonard Bernstein's Mahler Ninth score had been viewed nearly 25,000 times.
- The project has been covered in *The New York Times*, *The Wall Street Journal*, *MusicalAmerica.com*, *Ariama.com*, *The Rest is Noise* (*New Yorker* critic Alex Ross's blog), *Playbill.com*, *WQXR.org*, and other local, national, and international media outlets.

“ The Digital Archives Project...offers a scholar's dream treasure trove that is instantly accessible to anyone in the world with internet access. ”

— Daniel Eno, *Ariama.com*

What's Next

The Philharmonic has upgraded to Alfresco 3.4.1 and is continuing to digitize and ingest content into Alfresco. Over the next 10 years, the Philharmonic plans to digitize its entire collection of 8 million pages of documents and 7,000 hours of audio visual material, reflecting the Philharmonic's continued commitment to providing the broadest possible access to its collections. When finished, the repository is expected to contain more than 2 petabytes of data and 160 years of archival information available for instant retrieval. The Philharmonic plans to develop partnerships with academic institutions and music conservatories to create curriculum that will focus on material available in the Digital Archives.

NY Philharmonic is also looking to implement Alfresco's Activiti BPM platform, a light-weight workflow and Business Process Management tool to further streamline the content approval process.

“ Perhaps the most intriguing news...can be found in the orchestra's new, fully searchable digital archive...comprehensive beyond anyone's wildest dream. ”

— Susan Elliott, *MusicalAmerica.com*

“ Alfresco is the ideal content platform to handle the sheer scope and scale of the New York Philharmonic Digital Archives project. It provides a modern architecture with the flexibility customers need to manage massive amounts of data — whether it is for historical or business content. We are pleased to be part of this important undertaking to make valuable historic content available for everyone around the world to view and listen to. ”

— Dave Giordano, Founder and President at Technology Services Group, Inc.