Repository solutions for time-based media
Panel presentation proposal, Open Repositories 2013

Archiving and delivering time-based media presents a number of special challenges for repository development and management. Workflows and storage systems must be scaled to adequately handle much larger files, metadata schema and tools must be flexible enough to describe content with complex internal structure, and multiple points of integration are needed to securely connect to streaming servers, authentication services, course management systems and other specialized services. The panelists will discuss efforts underway at their institutions to develop repository solutions for time-based media management and delivery.

Karen Cariani, WGBH
WGBH has been awarded an NEH (US) Preservation and Access Research and Development grant to build an open source digital asset management system for media preservation. WGBH needs flexible systems that can easily evolve, as the broadcast industry is constantly changing with vendors creating new cameras, editing systems, and other hardware and software. The repository solution must be flexible enough to accommodate basic format changes in the digital content source, to accommodate content and file complexity and flexibility in adding metadata and content descriptions.

Declan Fleming, University of California San Diego
The UCSD DAMS (https://libraries.ucsd.edu/digital/) takes an object approach to digital assets. Recognizing that different types of objects require varying levels of metadata, we adopted an RDF triplestore model to describing objects, leveraging as many existing standards as possible and creating triples for specific needs when necessary. The DAMS houses any kind of object and metadata, including audio and video, both as single objects and complex objects containing links to subcomponents with metadata. In the past, we have made these objects accessible for download and have recently begun the integration of a Wowza streaming server to enhance the user experience. Development of the DAMS has required significant attention to a robust storage infrastructure, and has laid important ground work for management of research data.

Todd Grappone, University of California Los Angeles
The UCLA Library NewsScape Library of International Television News (newsscape.library.ucla.edu) is a large and growing digital collection of over 195,000 news programs recorded from dozens of major television networks from 2005 to the present. Initiated by faculty in the UCLA Department of Communications Studies, the collection also includes more than 1.1 billion words of closed caption texts and transcripts and over 44 million thumbnail images. These materials facilitate full-text searching and visual navigation of the archive's contents via a Web-based interface (Solr), which provides instant streaming access to the recorded news programs. The archive is stored at the UCLA Library, where it occupies 47 terabytes of disk storage and grows at a rate of nearly 800 gigabytes per month, adding

approximately 110 hours of recorded television each day. A separate, interdisciplinary research group, headed by Professor Steen and faculty members from the fields of Statistics and Computer Science, has received a grant from the National Science Foundation to develop new means of exploring the archive, including sophisticated text mining of captions and transcripts and also automated face detection and on-screen text recognition. These capabilities will soon be merged into the Library's service alongside tools that will allow the archive's contents to be incorporated into scholarly publications and undergraduate curricula. The NewsScape archive is currently available on UCLA campus only with plans to launch access to the rest of the UC coming in the fall as we develop policy for and comfort with associated copyright issues.

Claire Stewart, Northwestern University
The Avalon Media System is an IMLS (US) funded project to develop an open source digital audio and video management system for libraries and archives, building upon a number of existing technologies, including Opencast Matterhorn, Fedora Commons, Hydra/Blacklight, and Red5. Avalon is being jointly developed by Indiana University and Northwestern University with the participation of 10 additional partner institutions. Working with two existing open source communities, supporting a variety of authentication and authorization approaches, and designing a system flexible enough to accommodate very different approaches to capture and description have been significant challenges. Adopters expect sufficient flexibility to permit integration with other media services and systems, such as Kaltura and Adobe Media Server. This session will offer a brief overview of the primary use cases and the strategies the distributed development team uses to iterate through the complex needs of the system.

Brian Tingle, California Digital Library
Calisphere is a free, public website of over 230,000 primary sources such as historical photographs, letters, drawings, and other artifacts, with a particular emphasis on California's storied history. These materials have been digitized and contributed by more than 115 libraries, archives, and museums throughout the state. The result is a veritable treasure trove of materials from California's most venerable institutions, free and open to online exploration from anywhere in the world. Calisphere added support for a small number of digitized audio and video recordings early in 2013. Rather than setting up and maintaining a streaming media server such as Wowza locally at CDL, time based media files are being hosted in Amazon Web Services using the CloudFront Content Distribution Network (CDN). Media files placed into AWS Simple Storage Service (S3) "buckets" are enabled in the CDN for both Adobe's Real Time Messaging Protocol (RTMP) streaming and http progressive download. Using a CDN enables popular content to be cached in "edge" locations worldwide. The MediaElement.js library is being used so that the HTML5 markup for audio and video elements is supported across all modern browsers with a consistent look and Google Analytics integration. The presentation will discuss encoding and presentation challenges addressed by our cross browser fallback strategy for media display and present data on the use of the materials and the costs associated with delivering content from the cloud.