# Media Preservation and Access with HydraDAM2 and Fedora 4

Karen Cariani
WGBH Media Library and Archives

Jon W. Dunn
Indiana University Bloomington Libraries

## Introduction

WGBH Media Library and Archives and the Indiana University Libraries have embarked on a two year project to extend the WGBH's existing HydraDAM digital asset management system to be able to serve as a digital preservation repository for time-based media collections at a wide range of institutions using multiple storage strategies. This new system will be based on the open source Hydra repository application framework and will utilize the emerging Fedora 4.0 digital repository architecture.

Audio and video files are significantly bigger than most other files that libraries and archives currently manage in digital repository systems, and as a result they need to be stored and handled differently. The management, discoverability, and preservation of large files and complex objects is difficult. Fedora 4, the latest iteration of the Fedora repository system, is built on a new architecture that promises to have capabilities to better support management and preservation of large files.

## Project Goals

1. Extend the HydraDAM digital asset management system to operate on Fedora 4
2. Develop Fedora 4 content models for audio and video preservation objects, including descriptive, structural, and digital provenance metadata, based on current standards and best practices and utilizing new features in Fedora 4 for storage and indexing of RDF
3. Implement support in HydraDAM for two different storage models, appropriate to different types of institutions:
    - direct management of media files stored on spinning disk or on tape in a hierarchical storage management (HSM) system
    - indirect management and tracking of media files stored offline on LTO tapes
4. Integrate HydraDAM into preservation workflows that feed access systems at IU (Avalon) and WGBH (Open Vault) and conduct testing of large files and high-throughput workflows
5. Document and disseminate information about our implementation and experience to the library, archive, digital repository, and audiovisual preservation communities

## WGBH Use Case

WGBH faces two core issues with the preservation of media files: 1) the size of preservation media files (being very large and thus processing time long), and 2) the large number of incoming born digital files that may be small but are many.

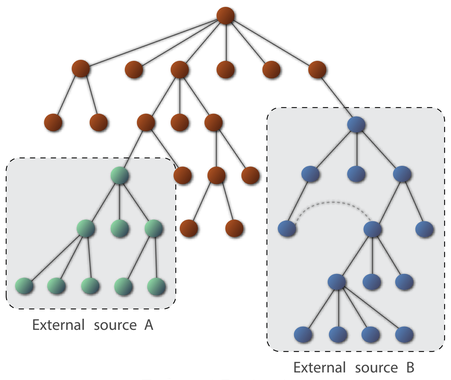
With the original HydraDAM1 application, functionality is based on the workflow of self-deposit of digital files and self-describing of metadata. This workflow is sufficient to ingest a small number of files to your repository, or to ingest files without any previously created descriptive records. The original HydraDAM1 wasn't designed to easily ingest batches of already created metadata and attach the records to many digital files.

WGBH has a 20 year old rich descriptive metadata database that is updated daily. Additional metadata lives in a robust Fedora 3 repository powering our Open Vault website. Adding the multiple descriptive data records into HydraDAM1 while ingesting many digital files would be inefficient as it is currently designed.

With HydraDAM2, we want to be able to ingest already existing metadata records into the repository, ingest location of proxy files and essence files (either on spinning disk or LTO tape stored in the vault) and attach the data to multiple files. Also, our users want to be able to search in a system to discover the digital assets we have available via proxy. Blacklight and Solr gives us that functionality.

## Key Fedora 4 Features: Existing and Planned

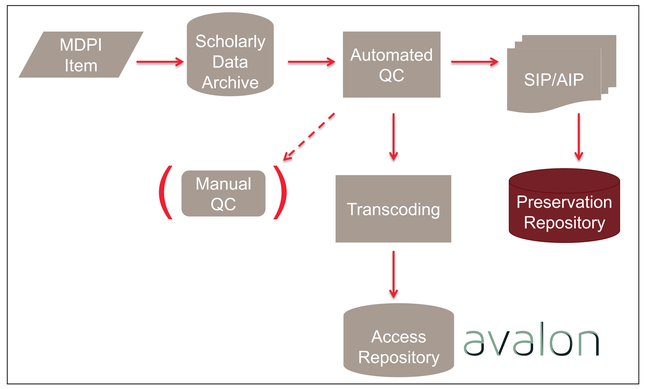
- Content modeling
    - Relationships
    - Metadata
- Linked data
    - PBCore/EBUCore
- Large file support
- Federation/projection
    - Fedora management of data in HSM




Fedora 4 Projection

## IU Use Case

Indiana University has embarked on ambitious plan to digitize all audio and video objects on its campuses judged to be important by experts by the time of the University's bicentennial in 2020. The collections to be digitized include nearly 300,000 objects held by over 80 campus units. Preservation-level digitization will be conducted in partnership with a private vendor and is scheduled to begin in spring 2015, producing nearly 7 petabytes of audio and video preservation files over a 3.5 year time period. HydraDAM2 will be used to manage audio and video preservation files and metadata stored in IU's HPSS-based hierarchical storage management system.




## Path to Fedora 4

The current version of HydraDAM is built on version 5 of Sufia, the self-deposit institutional repository Hydra Head originally created by Penn State. The HydraDAM2 team plans to take advantage of Sufia's support for Fedora 4 (starting in version 6) and upcoming support for the new Portland Common Data Model.

Fedora™ hydra

INDIANA UNIVERSITY WGBH NATIONAL ENDOWMENT FOR THE Humanities