

## **File and Directory Naming Conventions and Workflow for Large Objects**

**The /rarch Directory on mss3.** As discussed previously for the R5.0 release, we will need a directory tree on mss3 to accommodate the archival masters. From previous discussions, we have decided that all video archival masters will use the “R” control group and will be located in a separate directory as part of the permanent object store. The motivation is to avoid segmenting large files and to improve ingest times. It is proposed that this web-accessible directory be set up as /rarch with an estimated size of 1TB. There was much discussion about the naming conventions to use for the /rarch subdirectories.

**Video Object Architecture.** The video object architecture is shown in Figure 1 at the end of this document. Note that two datastreams (SMOV-1 and XACML) have been added from the previous version. The SMOV-1 datastream is a QuickTime stream which will be ingested and preserved as a managed file. This file will also be duplicated on the Darwin video server (see notes below). Secondly, for NJVid licensed videos, there will be an XACML policy. Selecting and ingesting XACML policies is part of our release R5.1 and is not discussed in this document. Per the xacml specification, we do allow an xacml policy to be added manually in R5.0 to accommodate an ETD embargo.

**Procedures for R5.0.** The objective is to establish file and directory naming conventions for R5.0 and subsequent releases. We will also want to use these conventions for the ingest of videos into R4.5.

*The Structure of /rarch.* To simplify directory structures and to create an efficient process, we decided on the following structure for the /rarch directory (refer to the Pre-ingest Directory structure below):

- The /rarch directory will be initially created in the Fedora object store area of mss3.
- All files to be ingested with the “M” attribute or to be referenced with “R” will be manually loaded into object level directories in /rarch. These files include the archival master, Flash, QuickTime, and the SMIL file. See Pre-ingest Directory below for an example for a typical video object.

### *WMS Procedures for R5.0*

- WMS will reserve the Fedora PID object which will be used to rename the object level directories and the archival master in the /rarch directory.
- WMS will construct the “R” url pointer to the archival master as follows:  
server/rarch/\$PID/\$PID+RARCH1.tar (e.g. server/rarch/rutgers-lib\_20000/rutgers-

lib\_20000+RARCH1.tar). This url is inserted in the RARCH1 datastream of the video object.

- WMS creates the SMIL XML file by copying the SMIL template, located in a WMS directory, and adding the url pointer that references the video stream on Darwin to the SMIL text file. The pointer would be of the form: rtsp://server/(Darwin file system)/rutgers-lib\_20000/rutgers-lib\_20000+SMOV-1.mov. The resulting SMIL file becomes a managed “M” datstream (similar to SMAP1) with ID=MOV-1. Note, in order to be compatible with the current version of WMS, the SMIL file must use a file type of .mov. This approach has been tested and works on lefty.
- WMS will upload the Flash (.flv, ID=FLV-1) and video stream QuickTime (.mov, ID=SMOV-1) files as per the normal procedure (except these files are in a different working directory, i.e. /rarch).
- WMS initiates a Fedora ingest of the object.
- As a post-ingest action, WMS renames the object directory and associated files. WMS also initiates an FTP file transfer of the QuickTime stream (ID=SMOV-1) to Darwin.
- At some later date, per our existing workarea policy, all files in the object level directory are deleted EXCEPT for the archival master. This results in the Post-ingest Directory Structure as shown below.

### **Workflow for R4.5**

When ingesting videos for NJVid, we will use the normal procedures for ingesting content, i.e. there will be no changes in WMS or any other software in R4.5 to accommodate the ingest of the videos. The objective will be to take the necessary manual steps so that the NJVid collections are compatible with the file and directory structures outlined above as part of R5.0. The following workflow is required:

- The files to be ingested are manually located in the WMS workarea with file names that can be properly associated with the ingested object, e.g. title1.tar, title1.mov, etc.
- The files in the workarea with Fedora datastream IDs, per the object architecture, are as follows:
  - Flash (.flv, ID=FLV-1)
  - SMIL text file (.mov, ID=MOV-1) with default url to be replaced later.
  - The QuickTime video stream (.mov, ID=SMOV-1)
  - A “dummy” file (.tar, ID=ARCH1) representing the archival master.

- The archival master file (.tar, ID=RARCH1) to be later renamed and relocated to /rarch
- In WMS, metadata is prepared and the Flash, QT, SMIL, and dummy files are ingested.
- Upon successful ingest, the title of the video and the Fedora PID are forwarded to someone with dlr/EDIT access. The following edits and actions are executed:
  - The files in the workarea are renamed using the Fedora PID.
  - The archival master is copied to /rarch using the same naming convention as outlined above, i.e. /rarch/rutgers-lib\_20010/rutgers-lib\_20010+RARCH1.tar
  - The QT stream is ftp-ed to Darwin.
  - The url in the SMIL file is updated to point to the QT stream on Darwin
- At this juncture, the object architecture has both an ARCH1 and an RARCH1. In R4.5, the signature checker will continue to verify the dummy ARCH1 and ignore RARCH1.
- After release of R5.0, the following should be done to provide valid signatures:
  - The dummy archival master (ARCH1) is purged.
  - The TechMD with dummy signature is edited to reflect the signature of the valid archival master in /rarch. Note that R5.0 signature processing will be able to verify signatures for archival masters in /rarch.

After these steps have been completed, the video object has the structure as shown in Figure 1, without the PDF and XACML streams, and the directory structure as shown in the Post-ingest Directory Structure below. These objects and the /rarch directory of archival masters will be compatible with R5.0

### **The Pre-ingest Directory Structure**

#### **/rarch (located on mss3 Fedora object area)**

/title1

title1.tar	(archival master, repository ID=RARCH1)
title1.mov	(A text SMIL file, repository ID=MOV-1)
title1s.mov	(QT video stream, repository ID=SMOV-1)
title1.flv	(Flash file, repository ID=FLV-1)

/title2

title2.tar	(archival master, repository ID=RARCH1)
title2.txt	(Smil file, repository ID=MOV-1)
title2s.mov	(QT video stream, repository ID=SMOV-1)

title2.flv (Flash file, repository ID=FLV-1)

### The Post-ingest Directory Structures for mss3 and Darwin

#### /rarch (located on mss3 Fedora object area)

/rutgers-lib\_20000

rutgers-lib\_20000+RARCH1.tar (archival master, repository ID=RARCH1)

/rutgers-lib\_20010

rutgers-lib\_20010+RARCH1.tar (archival master, repository ID=RARCH1)

...

#### /(Darwin File System)

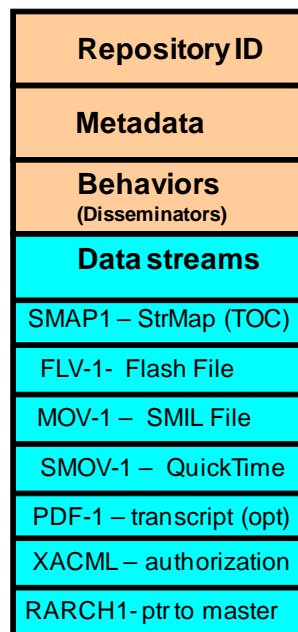
...

/rutgers-lib\_20000

rutgers-lib\_20000+SMOV-1.mov

/rutgers-lib\_20010

rutgers-lib\_20010+SMOV-1.mov



**Figure 1 – Video Object Architecture for R5.0**

**Summary**

After several review, this document was approved in the August 14 Working Group meeting. We agreed to review the ID naming conventions for MOV-1 and SMOV-1 for the R5.0 release. It was noted that we need to follow these conventions for R4.5 and the ingest of videos for NJVid. However, we might want to move to a convention in R5.0 in which the SMIL datastream takes on the a SMIL-1 ID and the QT stream is changed from SMOV-1 to MOV-1. If we make these changes, we would need to edit the 30 or so videos that are ingest in R4.5.

rcj – 08/22/2008