

OR 2014
Helsinki

Bringing Together Resources Using Fedora 4

Goals

- Explore and Test Fedora 4 with real data
 - consolidate repository
 - re-think data organization
 - support ongoing curation
 - start dialog

Local Broadcast News Media


“Entertaining and informing with
sensational headlines and dubious
reporting for over 50 years”



WSLS Television News Collection

- 13,000 16mm local news film clips
- 18,000 pages of anchor scripts
- NEH grant for digitization, preservation and access

Collection Navigation

 **WSLS-TV (Roanoke, VA) News Film Collection, 1951 to 1971**

-  1951
-  1953
-  1954
-  1955
-  1956
-  1957
-  1958
-  1959
-  1960
-  1961
-  1962
-  1963
-  1964
-  1965
-  1966
-  1967
-  1968
-  1969
-  1970
-  1971
-  unknown

Item Details (Item 3 of 435)

[← Return to search results](#) | [Start over](#) |  [Add Star](#)[Item Actions](#) ▼[« Previous](#) | [Next »](#)

WSLS-TV (Roanoke, VA) News Film Collection, 1951 to 1971

[Search Inside Collection](#)

In 2007, WSLS-TV of Roanoke, Va., gave news film and scripts from their mid-20th century broadcasts to the University of Virginia Library for preservation and use. The resulting collection spans 1951 to 1971 and comprises approximately 13,000 clips of 16mm film shown during news programs. It also contains roughly 18,000 pages of the accompanying scripts read on air by anchorpersons.

In 2010, the National Endowment for the Humanities awarded a grant of \$254,600 to the Library for the preservation and digitization of both the film and the paper scripts. The NEH designated the collection a part of its "We the People" initiative, which seeks to encourage and enhance the study and understanding of American history, culture, and democratic principles.

The WSLS-TV news film collection is now available to the public in digitized form through the Library's online catalog, Virgo. When a news clip and anchor script cover the same news story, the video and a PDF of the script will appear together in one catalog record. There are some news clips that do not have anchorperson scripts, and likewise, there are some scripts for which a matching video has not been found. These variations will be displayed in their own catalog records.

Content of the news stories varies greatly from soapbox derbies to civil rights demonstrations. Local affairs and politics, sports, community events and social occasions, and personalities of the Roanoke, Va., region constitute the bulk of the collection. Content also includes a significant amount of coverage of Massive Resistance and the Civil Rights Movement in Virginia. National figures and issues that affected the Roanoke area, such as visits from U. S.

A vintage wooden television set with a dark screen. The screen displays the text "Up Next..." in a bold, italicized font, followed by "Global Events" and "Local News" in a bold, sans-serif font. The television has a wooden frame and a control panel on the right side with two knobs and a speaker grille.

Up Next...
Global Events
Local News

PIDs missing, feared dead!

- UUIDs
- resolvable hierarchy-based URIs
- ...even namespaced:names (changeme:1)

Pro-Tip: namespaced names + apache mod_rewrite means you can keep your old URLs.

Parent-Child beauty pageant to be held.

- arbitrary relationships still allowed/encouraged
- hierarchy has special uses
 - relationship implied in URIs
 - versioning, deleting subgraph
 - inherited AuthZ policies

Armed RDF militants overthrow XML despot.

- ... dragged, kicking and screaming into the linked data future...
- ... XSLT retired... LDPath promoted...
- ... retirement home exists for relics of the past.

Solr Fuseki accused of wiretapping!

- jms-indexer-pluggable implicated in conspiracy...

First successful brain transplant!

- Federation allows existing resources to be grafted into the repository.
- Limitations apply, but potential uses are vast.

***Scientists develop time travel;
ponder whether to use it***

- as much versioning as you want, no more, no less.

A vintage wooden television set with a large screen and control knobs on the right side. The screen is dark and reflects some light. The text is centered on the screen.

**.. and in local
news..**

Data Modeling (Fedora 3 state of the art)



Fedora Digital Object
List Datastreams

Object Identifier (PID): uva-lib:2266782

Version Date: current

| Datastream ID | | Datastream Label | MIME Type |
|-----------------------------|------------------------------|------------------------------------|---------------------|
| DC | view content | Dublin Core Record for this object | text/xml |
| metadata | view content | PBCore metadata | text/xml |
| POLICY | view content | | text/xml |
| solrArchive | view content | Index Data for Posting to Solr | text/xml |
| RELS-EXT | view content | Relationships | application/rdf+xml |



Fedora Digital Object
List Datastreams

Object Identifier (PID): uva-lib:2266783

Version Date: current

| Datastream ID | | Datastream Label | MIME Type |
|---------------------------|------------------------------|------------------------------------|---------------------|
| DC | view content | Dublin Core Record for this object | text/xml |
| RELS-EXT | view content | Relationships | application/rdf+xml |
| scriptPDF | view content | Anchor Script (PDF) | application/pdf |
| scriptTXT | view content | Anchor Script (keyed text) | text/plain |
| thumbnail | view content | Thumbnail image of anchor script | image/png |



14,000

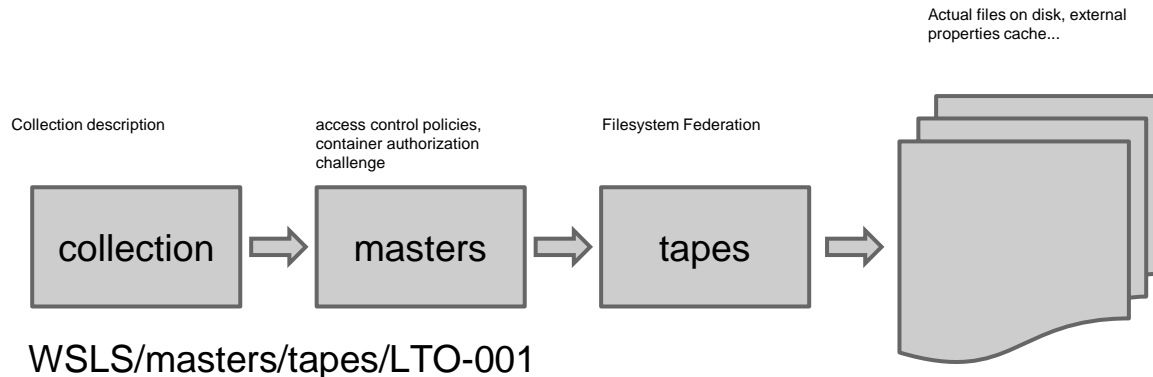
Data Modeling

(Fedora 4 remix)

- could port the existing model to fedora 4
- but can't we do better??
 - a. include master files
 - b. use the organization that is intuitive to those working on the collection (and those who aren't)
 - c. free ourselves from historic metadata compromises, and future metadata migrations

Preservation Master Video Files

- 500MB - 5GB individual files, ~12T total
- Preservation needed years before cataloging began
- access restrictions



FS Federation Findings

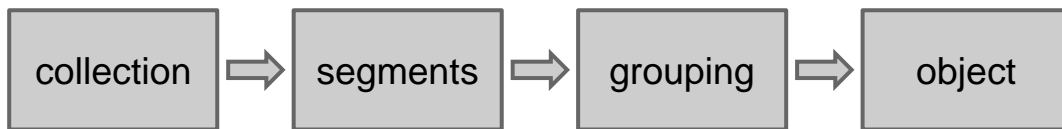
- Read-Only
- External properties cache
 - checksums
- Hide your mess with symlinks
- Avoid expensive calls until you need them
 - long file listings (triggers initial checksum computation)
 - add direct links from logical items

Metadata Records

- legacy collection-specific IDs
- born in spreadsheets
- produced through countless hours of trained labor
- grudgingly PBCore XML

Segment Objects

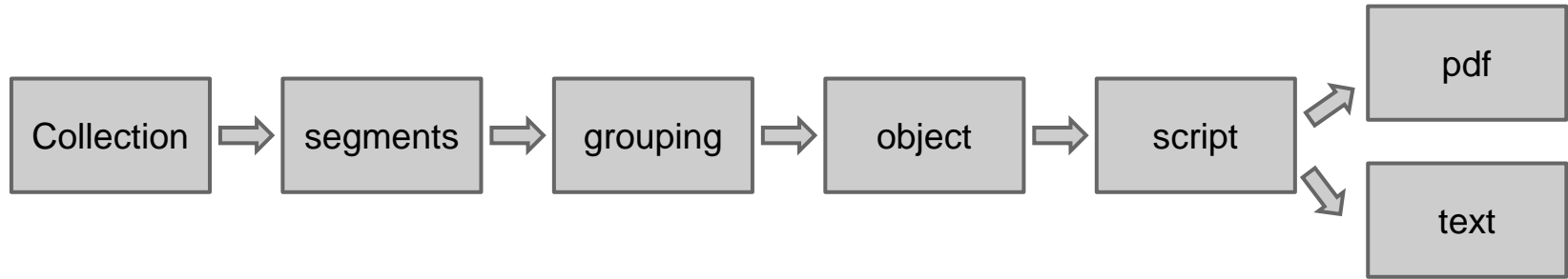
- Upgrade metadata RDF that is
 - indexed synchronously
 - cheaper to modify
 - can have types
 - can live alongside other assertions
 - is part of the object



WSLS/masters/segments/0000-0999/0003_1

Script Objects

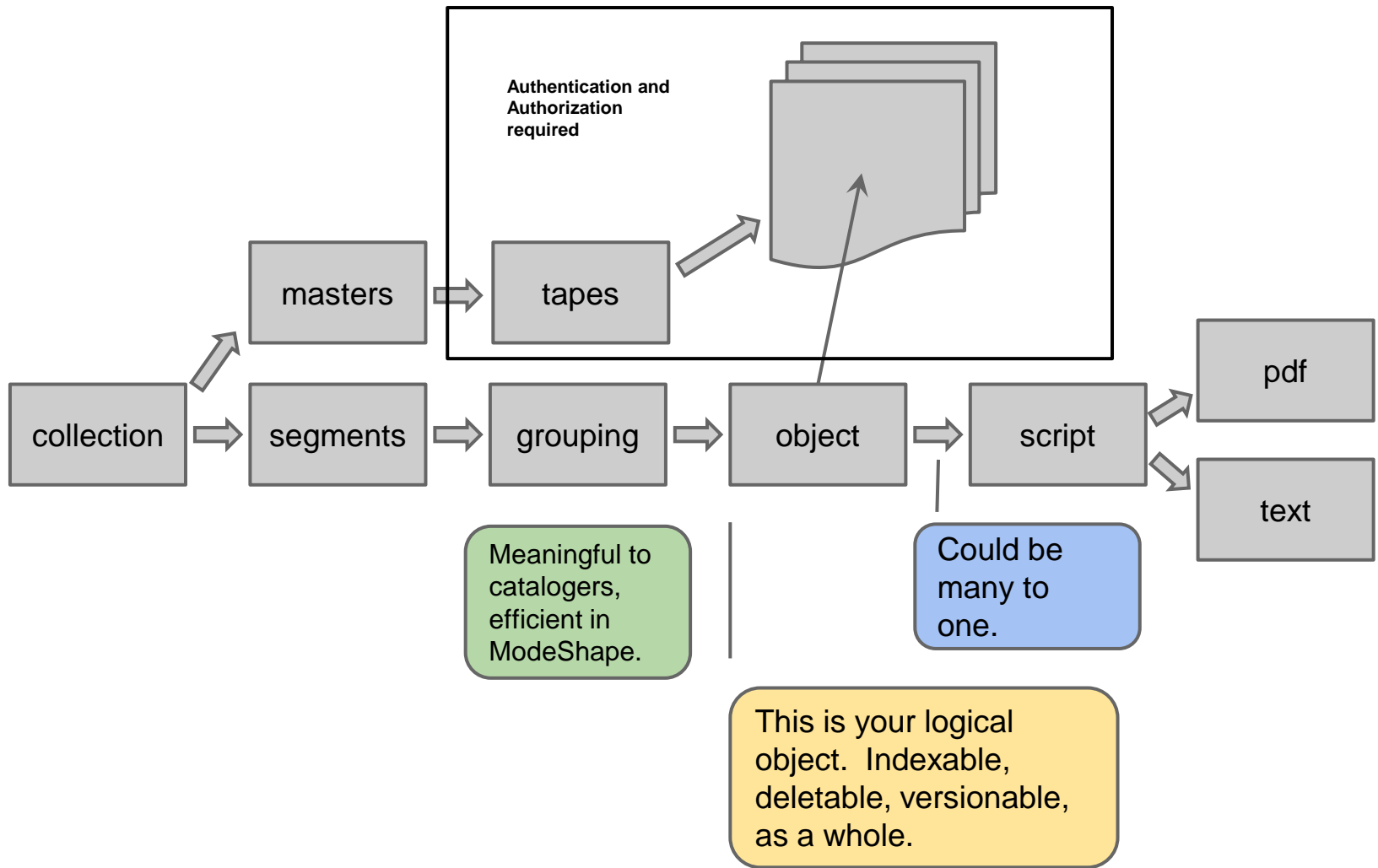
(and their datastreams)

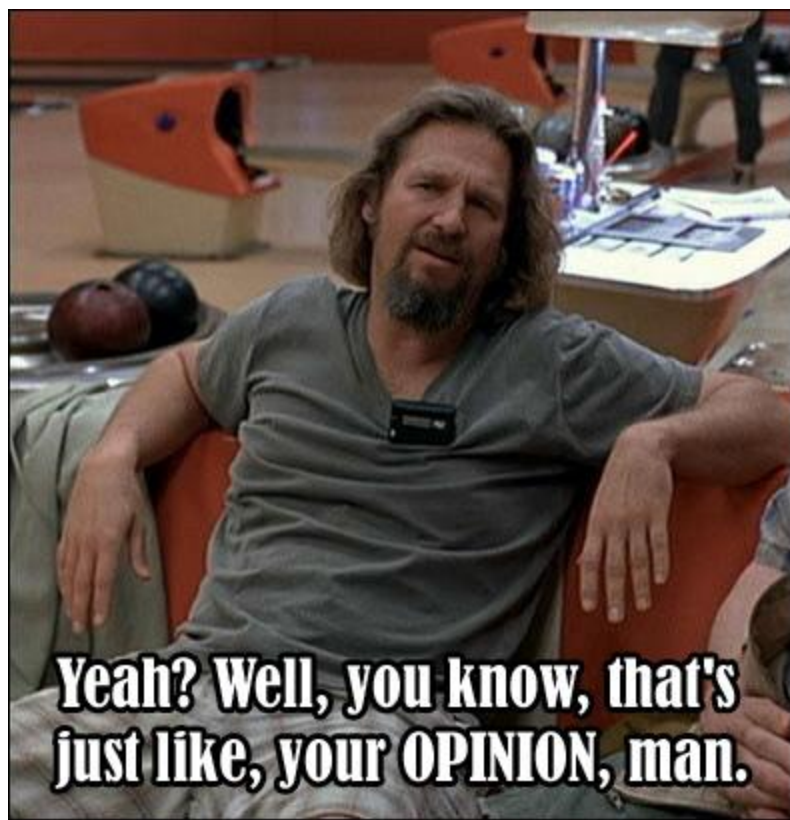


WSLS/masters/segments/0000-0999/0003_1/script

WSLS/masters/segments/0000-0999/0003_1/script/pdf

WSLS/masters/segments/0000-0999/0003_1/script/text





Pros

- It works*
- Allows secure inclusion of master videos at low cost
- Modeled after the organization and structure familiar to project participants

Cons

- some “clever” hacks required
- meaningful IDs will always burn you
- more testing needed

Big Dreams -- Unfinished Work

- indexing and deployment
- streaming file integration (avalon cough avalon)
- SSO integration, XACML
- automatic thumbnail generation

Thoughts?

md5wz at Virginia EDU
mikedurbin on Github

Credits

Television Photograph - CC - ellenm1 - <https://www.flickr.com/photos/ellenm1/>