

Interoperability of Metadata for Thematic Research Collections A Model Based on the Walt Whitman Archive

Mapping from METS to Fedora/UVa

Overview

Fedora is an open source digital library repository. Fedora employs a highly flexible architecture that supports any type of digital object, multiple views of each, and relationships among them. In principle, Fedora is technologically capable of accommodating the variety of digital objects comprising the Walt Whitman Archive.

Each repository implementing Fedora, however, selects the types of digital objects it will support, including the method or methods of rendering, based on need and economy. Each type of digital object requires development before it can be collected and disseminated. Particularly labor-intensive are "behaviors" or renderings associated with each digital object. Behaviors, subsequent to collection, must also be maintained, as they are dependent on formats, standards, and programs that change over time. Given the development and maintenance cost of accommodating individual digital objects and associated behaviors, repositories attempt to collect a large number of objects of a small number of types. The University of Virginia has chosen to support a small number of objects that follow strict encoding or representation protocols and very specific behaviors for these types.

Because of this policy, a large number of the object types employed in the Walt Whitman Archive (WWA) are not supported in Fedora as implemented by the University of Virginia (hereafter Fedora/UVa). Of the six forms objects identified (essays, archive, bibliography, books (and book-like objects, works, and portraits) books are the only type supported. Support for books-like objects, though, is limited to a specific TEI encoding (a subset of TEI Tite) and related page-images are assumed to be in at least two image sizes, a thumbnail image and screen-size. The policy of supporting this particular protocol with respect to book-like objects is based on economy, as developing and maintaining the behaviors for rendering the more complex encodings used in the WWA is labor-intensive.

While the Metadata Encoding and Transmission Standard (METS) is used by many implementers of Fedora, the University of Virginia uses an alternative though similar protocol, Fedora Object XML or Foxml. In

order to ingest WWA objects into Fedora/UVa, it is necessary to transform METS instances into Foxml instances. Given that WWA book-like objects similar to the book-like object implemented in Fedora/UVa, work at Virginia focused on mapping and transforming this object type.

The following objects were selected for transformation:

- | | |
|--|---|
| 1. mets.duk.00042-01.xml | Manuscript of "Poem of Materials" |
| 2. mets.med.00004-01.xml
4 (1953) | With Walt Whitman in Camden vol.
4 (1953) |
| 3. mets.per.00024-01.xml | Periodical printing of "Yonnondio" |
| 4. mets.per.00054-01.xml
of the Eagles" | Periodical printing of "The Dalliance
of the Eagles" |
| 5. mets.ppp.00473-01.xml | Leaves of Grass (1867) |
| 6. mets.ppp.00707-01.xml | Leaves of Grass (1891-92) |
| 7. mets.yal.00003-01.xml
Dakota" | Manuscript of "Italian Music in
Dakota" |

Of these, 2, 3, and 4 proved to be problematic. 2 was problematic because it was not a transcription, as such, and thus fell outside of Fedora/UVa protocol. 3 and 4 were problematic because the mets:structMap differed in encoding from 1, 5, 6, and 7. (See more detailed description below).

Fedora/UVa Book-like Object Requirements

Each Fedora/UVa book-like objects requires the following:

1. A Foxml instance for the TEI encoded instance and for each page image
2. Administrative metadata XML instance based on a locally defined DTD for the TEI encoded instance and for each page image
3. Descriptive metadata XML instance based on a locally defined DTD for the TEI encoded instance and for each page image

Each METS instance for a book-like object thus results in six types of files, three for each TEI instance, and three for each page image.

WWA METS to Fedora/UVa Mapping

Persistent Identifiers

All digital objects, including dependent digital objects such as page images, ingested into Fedora must have a persistent identifier (PID). Further, each descriptive metadata instance and each administrative metadata instance must as have a PID, distinct from the PID of the object itself. While each WWA METS instance has a unique identifier (mets:mets @OBJID), the dependent page images do not. Thus before mapping METS to Foxml (including additional descriptive and administrative metadata), an inventory of all objects to be collected was created using a simple ad hoc XML format. Using XSLT, each object was then assigned a PID. The PIDs assigned and the associated URL are used in the METS to Foxml transformations.

Sample entries, the first for a page image, the second for a TEI instance:

```
<object>
  <url>duk.00042\duk.00042.002.jpg</url>
  <pid>WWAPID:0094</pid>
  <desc>
    <url>duk.00042\duk.00042.002.desc.xml</url>
    <pid>WWAPID:0094D</pid>
  </desc>
  <admin>
    <url>duk.00042\duk.00042.002.admin.xml</url>
    <pid>WWAPID:0094A</pid>
  </admin>
</object>
<object>
  <url>duk.00042\duk.00042.xml</url>
  <pid>WWAPID:0095</pid>
  <desc>
    <url>duk.00042\duk.00042.desc.xml</url>
    <pid>WWAPID:0095D</pid>
  </desc>
  <admin>
    <url>duk.00042\duk.00042.admin.xml</url>
    <pid>WWAPID:0095A</pid>
```

```
</admin>
</object>
```

Access Restrictions

Because Fedora/UVa only supports restrictions on access based on UVa access policy, and the "permit-to-all " policy was roughly compatible with the rights expressed in the WWA METS instances, the mapping defaulted to accept the UVa policy. (Note for further discussion: this raises several important issues with respect to handling of rights and access to digital objects collected by libraries.)

METS to Foxml Mapping: TEI Instance

METS values given as XPath expressions.

1. `<foxml:property NAME="info:fedora/fedora-system:def/model#label" VALUE="{//mods:name[mods:role/mods:roleTerm='Author']/mods:namePart} : {//mods:titleInfo/mods:title}"/>`

Derived from the `<mods:name>` for the author and the `<mods:title>`

2. Dublin Core

- a. `<dc:creator><xsl:value-of select="//mods:name[mods:role/mods:roleTerm='Author']/mods:namePart"/>`
- b. `<dc:title><xsl:value-of select="//mods:titleInfo/mods:title"/>`
- c. `<dc:identifier>` PID from assigned list

3. `<foxml:datastream>`

- a. `<foxml:datastreamVersion>` @LABEL = "WWA" +
//mets:mets/@TYPE + " : " + URL of instance
- b. `<foxml:contentLocation>` @REF = URL of instance

METS to UVa Descriptive Metadata Mapping: TEI Instance

Known issues: for manuscript materials, source description resides in EAD instances, one for the WWA union guide and one for the holding repository. For printed books, no source information is given in the METS instance. The UVa descriptive metadata is a mix of description of

the digital instance and the original source. <surrogate> contains the description of the digital instance.

1. <pid> for descriptive data instance provided from assigned list
2. <mediatype type="text"><form scheme="WWA"> =
<xsl:value-of select="//mets:mets/@TYPE"/>
3. <presentation>
 - a. <displayAgent> = <xsl:value-of select="//mods:name [mods:role/mods:roleTerm='Author']/mods:namePart"/>
 - b. <sortAgent> = <xsl:value-of select="translate(//mods:name[mods:role/mods:roleTerm='Author']/mods:namePart,\$replace,\$with)"/>

Note: translate normalizes value to uppercase.

- c. <displayTitle> = <xsl:value-of select="//mods:titleInfo/mods:title"/>
- d. <sortTitle> = <xsl:value-of select="translate(//mods:mods/mods:titleInfo/mods:title,\$replace,\$with)"/>

Note: translate normalizes value to uppercase.

4. <agent type="creator" form="persname" role="author"><name> =
<xsl:value-of select="//mods:name [mods:role/mods:roleTerm='Author']/mods:namePart"/>
5. <title type="main"><xsl:value-of select="//mods:titleInfo/mods:title"/>
6. <identifier type="Library of Congress Classification"> =
<xsl:value-of select="//mods:classification [@authority='lcc']"/>

Note: Only if present in the METS

7. <language> = <xsl:value-of select="//mods:languageTerm [@authority='iso639-2b']" />

Note: Only if present in the METS

8. <physdesc type="extent">[Unavailable: in EAD record that is not given]</physdesc>

9. <title type="main"> = <xsl:value-of select="//mods:titleInfo / mods:title"/>

10. <place><geogname>Lincoln, Nebraska

Note: Value supplied in XSLT

11. <agent type="provider" form="corpname" role="publisher">
<name>The Center for Digital Research in the Humanities

Note: Value supplied in XSLT

12. <agent type="provider" form="corpname" role="publisher">
<name>Walt Whitman Archive</name>

Note: Value supplied in XSLT

13. <time type="creation"> <date> = <xsl:value-of select="//mods:dateIssued" />

14. <rights type="copyright"> = copyright symbol + <xsl:value-of select="//mods:dateIssued"/> + [through complicated use of xsl:key and referencing the correct set of rights metadata] rights:RightsHolderName

15. <rights type="access"> = Publicly accessible

Note: Value supplied in XSLT

METS to UVa Administrative Metadata Mapping: TEI Instance

1. <pid> for administrative data instance provided from assigned list

2. <character encoding="UTF-8"/>

3. <mimetype>text/xml</mimetype>

4. <markup>

a. <encoding label="xml" version="1.0"/>

b. <base @label="TEI.2" type="DTD" version="P4"
@href="http://www.tei-c.org/Guidelines/P4/p4dtd.xml"

c. <localmodification @label="TEI.extensions.dtd" @href="[path to
Whitman.ext.dtd]"

- d. <localmodification> @label="TEI.extensions.ent" href="[path to whitman.ext.ent]"

Note: All values in Administrative data provided by XSLT

METS to Foxml Mapping: Page Image

1. <foxml:property NAME="info:fedora/fedora-system:def/model#label" VALUE="{//mods:name[mods:role/mods:roleTerm='Author']/mods:namePart} : {//mods:titleInfo/mods:title} : {@LABEL}"/>

Derived from the <mods:name> for the author, the <mods:title>, and the @LABEL on the mets:div for the page.

2. Dublin Core

- a. <dc:creator><xsl:value-of select="//mods:name[mods:role/mods:roleTerm='Author']/mods:namePart"/>
- b. <dc:title><xsl:value-of select="//mods:titleInfo/mods:title"/> + @LABEL on the mets:div for the page.
- c. <dc:identifier> PID from assigned list

3. <foxml:datastream>

- a. <foxml:datastreamVersion> @LABEL = "WWA WWA Screen sized image:" + URL of page image file
- b. <foxml:contentLocation> @REF = URL of page image file

4. <foxml:datastream>

- a. <foxml:datastreamVersion> @LABEL = "WWA Thumbnail sized image:" + URL of page image file, modified for thumbnail size
- b. <foxml:contentLocation> @REF = URL of page image file, modified for thumbnail size

METS to UVa Descriptive Metadata Mapping: Page Image

1. <pid> for descriptive data instance provided from assigned list

2. <mediatype type="image"/>

Note: Value supplied by XSLT

3. <identifier type="parent"> PID value assigned to the TEI

4. <rights type="copyright"> = copyright symbol + <xsl:value-of select="//mods:dateIssued"/> + [through complicated use of xsl:key and referencing the correct set of rights metadata] rights:RightsHolderName

5. <rights type="access"> = Publicly accessible

6. <title><xsl:value-of select="//mods:titleInfo/mods:title"/> + @LEVEL on mets:div for page

METS to UVa Administrative Metadata Mapping: Page Image

1. <pid> for administrative data instance provided from assigned list

2. <identifier type="WWA FILE ID"> = file name from mets:FLocat @xlink:href

3. <adminrights><policy><access> = public

Note: Value supplied by XSLT

4. <technical><image>

a. <filesize type="kilobytes"> = value mets:file[@ID=\$fileID]/@SIZE, converted to kilobytes and rounded

b. <mimetype> = value from mets:file/@MIMETYPE

c. <compression> = Not specified

d. <colorspace> = Not specified

e. <imagewidth type="pixels">Not specified

f. <imagelength type="pixels">Not specified

g. <sourceX>Not specified

h. <sourceY>Not specified

Note: All values for c.-h. not available in METS. It is unclear how the values are used in Fedora/UVa, or if they are only provided for archival purposes.

TEI and Page Image Modification

In addition transforming the METS instance into one or more sets of Foxml/descriptive metadata/administrative metadata instances, it is also necessary to transform the TEI instance for a manuscript or print transcription, specifically to facilitate linking to page images as performed by disseminators for book-like objects in Fedora/UVa. Further, thumbnail images are used extensively with the Fedora/UVa implementation, and thus it is necessary to create thumbnail image version of all page images.