

# XML and Digital Libraries

# XML and the DL

- What are the possible uses of XML in a digital library?
  - Data storage / archiving
  - Data exchange / ingestion
  - Presentation / publishing
  - Metadata

# METS

- Metadata Encoding and Transmission Standard
- Attempts to solve the complex object problem
- Five components:
  1. descriptive metadata
  2. administrative metadata
  3. file groups
  4. structural map
  5. behaviors

# METS+

- METS doesn't, in itself, specify a metadata vocabulary, so what should you use?
  - MODS
  - Dublin Core
  - TEIHeader
  - ...

# Schemas and NS

- The first interesting problem:
- Our METS document needs to support multiple namespaces. How do we validate it?
- How do we attach the schemas to it?

# simple case

## No Namespaces:

```
<root xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:noNamespaceSchemaLocation="http://my.schema.example.com/  
schema.xsd">
```

## Namespaces:

```
<root xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xmlns="http://my.schema.example.com/ns"  
xsi:schemaLocation="http://my.schema.example.com/schema.xsd">
```

# multiple schemas

```
<mets xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"  
xmlns:dct="http://purl.org/dc/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/"  
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd  
http://purl.org/dc/elements/1.1/ http://uk.dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd  
http://purl.org/dc/terms/ http://uk.dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd">
```

# implementing METS

## CSR document table:

```
<row>
  <field name="doc_id">csr01-0001</field>
  <field name="volume">01</field>
  <field name="page_start">5</field>
  <field name="page_end">5</field>
  <field name="csr_title">The obseruations of Master John Pory Secretarie of Virginia in his trauels 1622</field>
  <field name="csr_source">Reprinted from Smith's History of Virginia, Vol. 2, P. 64.</field>
  <field name="doc_class">Histories</field>
  <field name="begin_date" xsi:nil="true" />
  <field name="begin_month" xsi:nil="true" />
  <field name="begin_year">1624</field>
  <field name="end_date" xsi:nil="true" />
  <field name="end_month" xsi:nil="true" />
  <field name="end_year" xsi:nil="true" />
  <field name="related" xsi:nil="true" />
  <field name="instructions" xsi:nil="true" />
  <field name="notes" xsi:nil="true" />
  <field name="docessay">document</field>
  <field name="time_stamp">2007-10-19 10:05:33</field>
  <field name="separation">0</field>
  <field name="problem">1</field>
  <field name="research" xsi:nil="true" />
  <field name="coverage">1622</field>
  <field name="new_title">Narrative of John Pory's travels [Extract]</field>
</row>
```

# implementing METS

## CSR Creator Table:

```
<row>  
  <field name="creator_id">10033</field>  
  <field name="creator_middle" xsi:nil="true" />  
  <field name="creator_dates">1572-1636</field>  
  <field name="creator_titles" xsi:nil="true" />  
  <field name="creator_pseud" xsi:nil="true" />  
  <field name="creator_suffix" xsi:nil="true" />  
  <field name="creator_source" xsi:nil="true" />  
  <field name="creator_qualifier" xsi:nil="true" />  
  <field name="creator_first">John</field>  
  <field name="creator_last">Pory</field>  
  <field name="type">person</field>  
</row>
```

# implementing METS

```
<dmdSec CREATED="2007-10-19T10:05:33" ID="csr01-0001">
  <mdWrap MDTYPE="DC">
    <xmlData>
      <dc:identifier>csr01-0001</dc:identifier>
      <dc:title>Narrative of John Pory's travels [Extract]</dc:title>
      <dc:subject>Histories</dc:subject>
      <dc:source>Reprinted from Smith's History of Virginia, Vol. 2, P. 64.</dc:source>
      <dct:extent>5</dct:extent>
      <dc:date>1624</dc:date>
      <dct:temporal>1622</dct:temporal>
      <dc:creator>John Pory</dc:creator>
    </xmlData>
  </mdWrap>
</dmdSec>
```

# implementing METS

```
<fileSec>
  <fileGrp>
    <file MIMETYPE="text/xml" ID="volume_01">
      <FLocat xlink:type="simple" xlink:href="volume_01.xml" LOCTYPE="URL"/>
    </file>
  </fileGrp>
</fileSec>
<structMap>
  <div TYPE="file">
    <div DMDID="csr01-0001">
      <fptr>
        <area BETYPE="IDREF" END="csr01-0001" BEGIN="csr01-0001" FILEID="volume_01"/>
      </fptr>
    </div>
  </div>
  ...

```