

# Greenstone Digital Libraries: Installation to Production Chopin Early Editions

Tod A. Olson <tod@uchicago.edu>

2005 ALA Annual Conference, Chicago IL

## Build overview

Preservation staff scan the physical scores and record structural metadata. Archival TIFFs are created, as are JPEG derivatives for web delivery. Detailed catalog records are created for the physical scores. These data are assembled into METS objects for each score, with MODS used for descriptive metadata. XSLT provides a custom transformation of METS to Greenstone Archive format, which is then loaded directly into the Greenstone software, which provides the indexing and user interface.

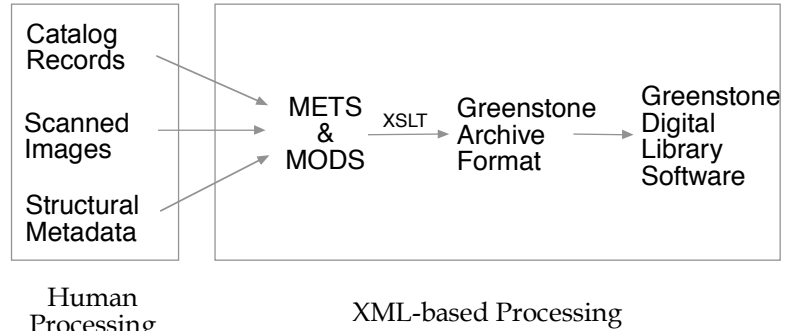


Fig 1: Build overview

## METS & MODS

The Metadata Encoding and Transmission Standard (METS) provides a general XML-based representation of digital objects, including: METS defines six sections: METS header, for information about the file itself; administrative metadata, including technical and rights management metadata; descriptive metadata; a file list; a structure map; and specified behaviors for files (e.g., send file to a server for interpretation). The structure map is the heart of METS; it permits arbitrary hierarchies to model digital objects.

Chopin Early Editions (CEE) uses only the structure map, descriptive metadata, and file list file sections at this time.

The Metadata Object Description Schema (MODS) is a schema for encoding of descriptive metadata. MODS preserves MARC-like semantics in a format that is more suitable to digital library projects: the data hierarchy is explicit; there is no catalog card-oriented field and punctuation structure. MODS provides special support for electronic resources and is extensible.

The following two pages show (1) an example of a real, but pared down, METS file from CEE, (2) a partial MARC record for a score in CEE, and (3) the MODS record generated from that MARC record.

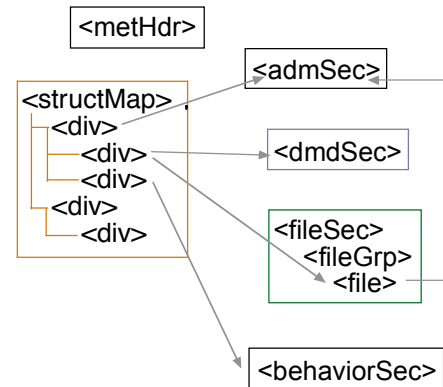


Fig 2: METS structure

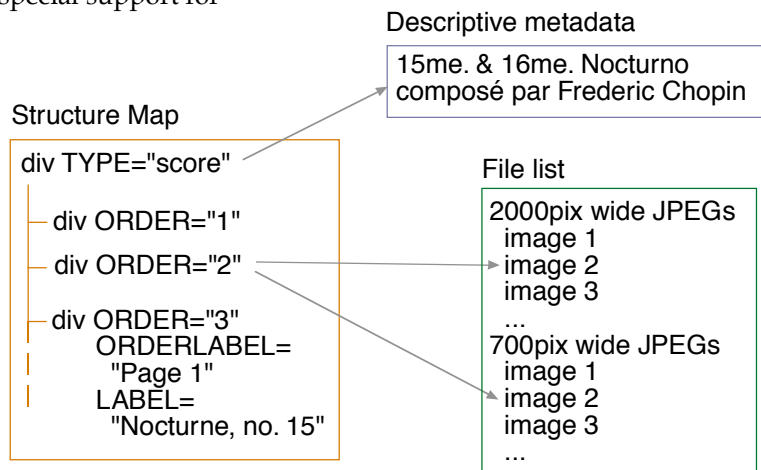


Fig 3: METS in Chopin Early Editions

# Greenstone Digital Libraries: Installation to Production

## Chopin Early Editions

### Sample METS Record

This is a reduced version of a METS record from CEE. Notice how the nesting of the <div> elements defines the structure of the score object, and how IDs link the structure to appropriate descriptive metadata and files. The top-level <div> represents the score object as a whole, and is linked to the descriptive metadata. The child <div> elements each represent a page of the score and link to image files. The order of the pages is explicit.

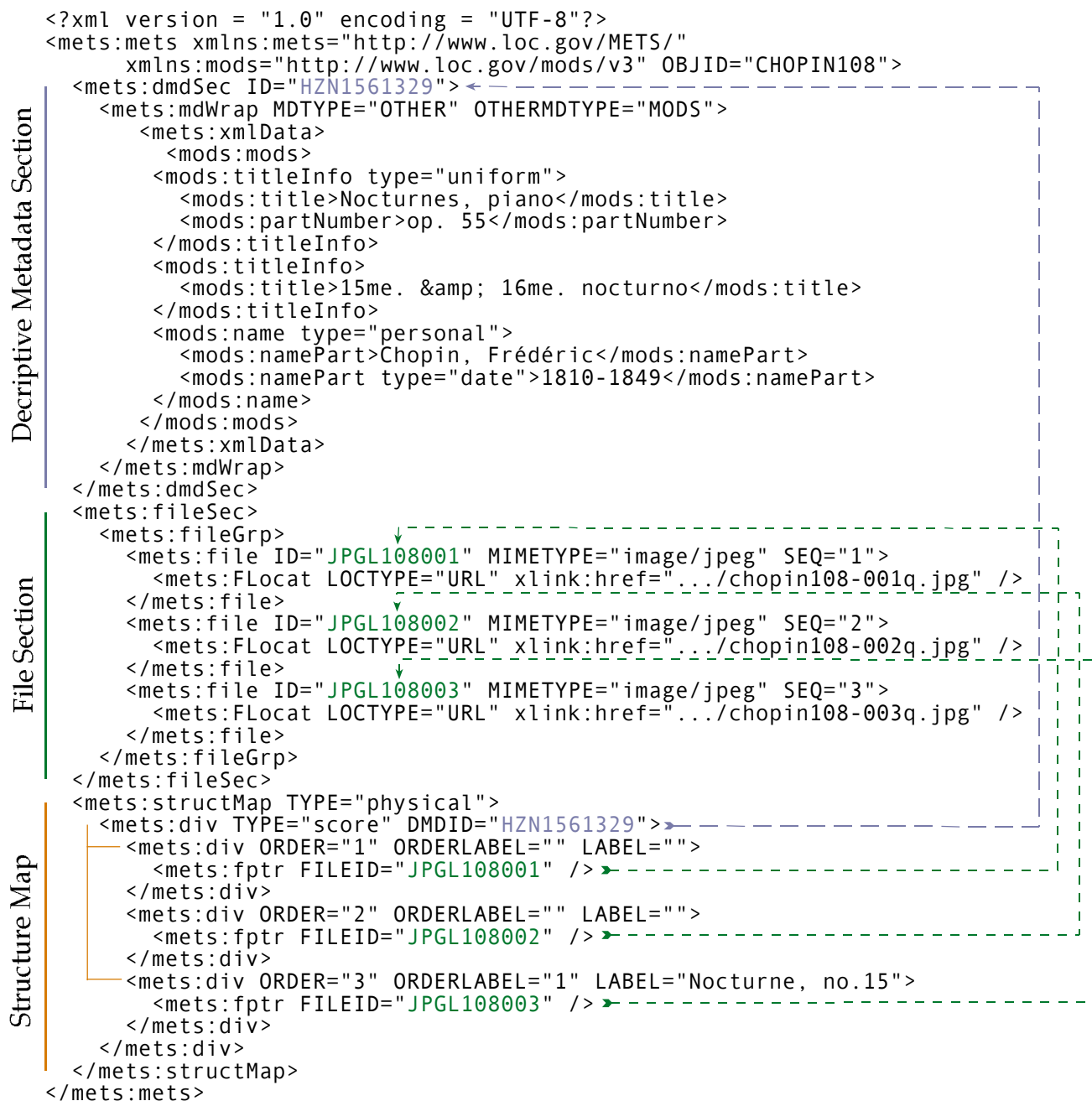


Fig 4: Sample METS record

## Greenstone Digital Libraries: Installation to Production Chopin Early Editions

### MARC and MODS samples

The MARC record is very concise, though the fields are cryptic. The data organization, including field order, data grouping, and embedded punctuation, is optimized for card printing. MODS prefers more intuitive tag names, an explicit hierarchical arrangement of data, and leaves the data formatting to the display logic.

```
100 1 $a Chopin, Frédéric, $d 1810-1849.
240 10 $a Nocturnes, $m piano, $n op. 55
245 10 $a 15me. & 16me. nocturno / $c [composé par Frederic Chopin].
246 3 $a Quinzième et seizième nocturno
260 $a London (No. 229, Regent Street, corner of Hanover Street) :
    $b Wessel & Co., importers and publishers of foreign music,
    $c [between 1848 and 1856]
650 0 $a Piano music.
800 1 $a Chopin, Frédéric, $d 1810-1849. $t Piano
    music (London, England) ; $v no. 59.
```

Fig 5: MARC sample

```
<mods>
<titleInfo type="uniform">
  <title>Nocturnes, piano</title>
  <partNumber>op. 55</partNumber>
</titleInfo>
<titleInfo>
  <title>15me. & 16me. nocturno</title>
</titleInfo>
<titleInfo type="alternative">
  <title>Quinzième et seizième nocturno</title>
</titleInfo>
<name type="personal">
  <namePart>Chopin, Frédéric</namePart>
  <namePart type="date">1810-1849</namePart>
</name>
<originInfo>
  <place>
    <placeTerm type="text">London (No. 229, Regent Street, corner of Hanover
Steet)</placeTerm>
  </place>
  <publisher>Wessel & Co., importers and publishers of foreign music</publisher>
  <dateIssued qualifier="inferred">between 1848 and 1856</dateIssued>
</originInfo>
<subject authority="lcsh">
  <topic>Piano music.</topic>
</subject>
<relatedItem type="series">
  <titleInfo>
    <title>Piano music (London, England)</title>
    <partNumber>no. 59</partNumber>
  </titleInfo>
  <name type="personal">
    <namePart>Chopin, Frédéric</namePart>
    <namePart type="date">1810-1849</namePart>
  </name>
</relatedItem>
<extension>
  <musicMd:genre
xmlns:musicMd="http://www.lib.uchicago.edu/XML/musicMd/">Nocturnes</musicMd:genre>
</extension>
</mods>
```

Fig 6: MODS sample, same content as in Fig 5

# Greenstone Digital Libraries: Installation to Production Chopin Early Editions

## Greenstone Archive Format (GSAF)

Documents in Greenstone are composed of Sections. Sections are made up of:

1. Description, which carries Metadata
2. Content for display
3. More Sections

Sections nest into a tree structure. Metadata and Content are at every level, and can be indexed and retrieved at the Section level. Content text may also be indexed down to the paragraph level.

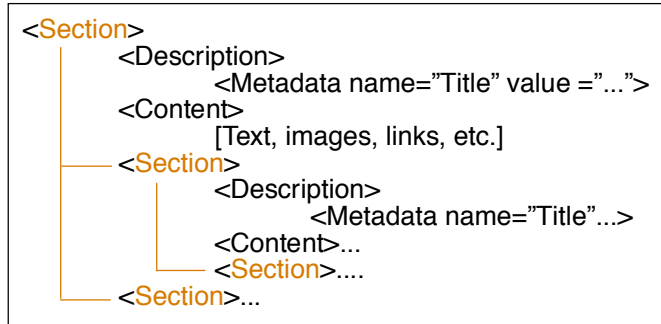


Fig 7: GSAF structure

## METS to GSAF

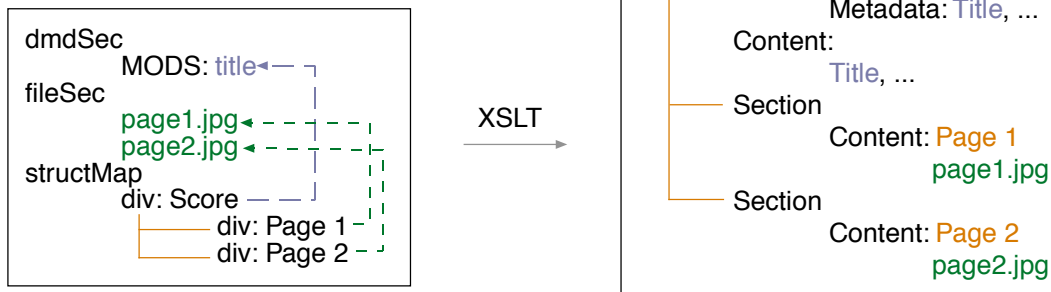


Fig 8: Transforming METS to GSAF

Both METS and GSAF represent an object as a hierarchy, so basic transformation is straightforward. The XSLT stylesheet walks the structMap and dereferences pointers to descriptive metadata and images. The MODS record is transformed into the metadata for searching and browsing the collection, and is inserted into the Description of the top-level Section; it is also transformed for display and placed in the top-level Content. Page-level references to the image URLs are transformed into HTML image tags and added to the page-level Content sections in the GSAF file.

## Custom navigation: page turner

Greenstone provides page turning by default, but a special variation was desired for CEE, shown in Fig 9a. The collection configuration file supports format strings, which dynamically generate HTML. Special metadata were generated in the GSAF files to add identifiers for the next and previous pages, as in Fig 9b. The DocumentText format string, shown in part in Fig 9c, uses custom metadata, together with Greenstone macros, to dynamically generate the links to the previous and next pages.



Fig 9a: custom page turner

```

<Section>
<Description>
<Metadata name="PrevPgId">CHOP108.1</Metadata>
<Metadata name="NextPgId">CHOP108.3</Metadata>
  
```

Fig 9b: GSAF metadata, next and previous section Ids

```

format DocumentText "... <table border=0... >
<tr>
  <td width=45%>
    {f}{[PrevPgId],
      <a href=\"_httpdocument_&cl=_cgiargcl_&d=[PrevPgId]\">
        _iconprev_ prev</a>,&nbsp;}</td>
  <td width=10%>&nbsp;</td>
  <td width=45%>
    {f}{[NextPgId],
      <a href=\"_httpdocument_&cl=_cgiargcl_&d=[NextPgId]\">
        next _iconnext_</a>,&nbsp;}</td>
  
```

Fig 9c: collect.cfg format strings operate on metadata and macros

# Greenstone Digital Libraries: Installation to Production

## Chopin Early Editions

### Custom navigation: document menu

Greenstone macros can be embedded in GSAF or in source documents, and they expand dynamically at runtime. Fig 10a shows content that generates a selection menu of all pages in a specific document. The values for the select options identify the desired sections of the document by combining the document ID, CHOP108, and a section number. This content, including macros, is expanded at runtime to the HTML shown in Fig 10b, which in turn is displayed by the browser as a drop-down menu, as shown in Fig 10c.

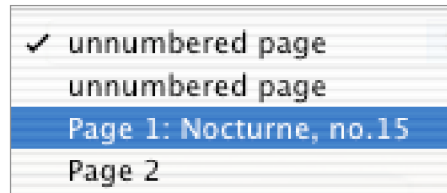
Notice that the document ID, CHOP108, is hardcoded in the values. This is inappropriate in many circumstances. If such a macro exists a better approach would be to use a GS macro that expands at runtime to the document ID.

```
<Content>...
<input type="hidden" name="e" value="_compressedoptions_" &gt;
<input type="hidden" name="a" value="d" &gt;
<input type="hidden" name="cl" value="_cgiargcl_" &gt;
<select name="d" onchange="submit()" &gt;
  <option value="CHOP108.1" selected &gt;unnumbered page</option&gt;
  <option value="CHOP108.2"&gt;unnumbered page</option&gt;
  <option value="CHOP108.3"&gt;Page 1: Nocturne, no.15</option&gt;
  <option value="CHOP108.4"&gt;Page 2</option&gt;
```

Fig 10a: GSAF content to generate menu

```
<form action="/gsdl/cgi-bin/library">
<input type="hidden" name="e" value="d-000-..." >
<input type="hidden" name="a" value="d" >
<input type="hidden" name="cl" value="CL3.11" >
<select name="d" onchange="submit()" >
  <option value="CHOP108.1" selected>unnumbered page</option>
  <option value="CHOP108.2">unnumbered page</option>
  <option value="CHOP108.3">Page 1: Nocturne, no.15</option>
  <option value="CHOP108.4">Page 2</option>
```

Fig 10b: resulting HTML for menu



✓ unnumbered page
unnumbered page
Page 1: Nocturne, no.15
Page 2

Fig 10c: page menu display

### Further Reading

*Chopin Early Editions*. <http://chopin.lib.uchicago.edu/>.

Cundiff, Morgan V. An Introduction to the Metadata Encoding and Transmission Standard (METS). *Library Hi Tech*, v. 22 no. 1, 2004, pp. 52-64. DOI 10.1108/07378830410524495.

*Greenstone*. <http://www.greenstone.org/>.

Guenther, Rebecca S. Using the Metadata Object Description Schema (MODS) for resource description: guidelines and applications. *Library Hi Tech*, v. 22 no. 1, 2004, pp. 89-98. DOI 10.1108/07378830410524521.

McCallum, Sally H. An Introduction to the Metadata Object Description Schema (MODS). *Library Hi Tech*, v. 22 no. 1, 2004, pp. 82-88. DOI 10.1108/07378830410524521.

*Metadata Encoding and Transmission Standard (METS)*. <http://www.loc.gov/standards/mets/>.

*Metadata Object Description Schema (MODS)*. <http://www.loc.gov/standards/mods/>.