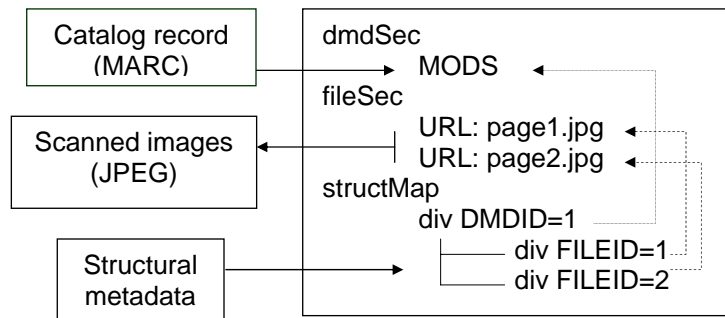# Building Collections Using Greenstone

Tod A. Olson <tod@uchicago.edu>
Sr. Programmer/Analyst
Digital Library Development Center
University of Chicago Library

METS & MODS

The catalog record for the score is crosswalked to MODS for the METS descriptive metadata. Page images are stored by naming convention, URLs are constructed from the structural metadata and added to the file inventory. The structure map has one top-level div for the score as a whole, which references the score's description. Each child div represents a score page, and refers to the page images in the file inventory.

```
Catalog record          dmdSec
(MARC)                        MODS
                        fileSec
Scanned images                URL: page1.jpg
(JPEG)                        URL: page2.jpg
                        structMap
Structural                    div DMDID=1
metadata                            div FILEID=1
                                    div FILEID=2
```
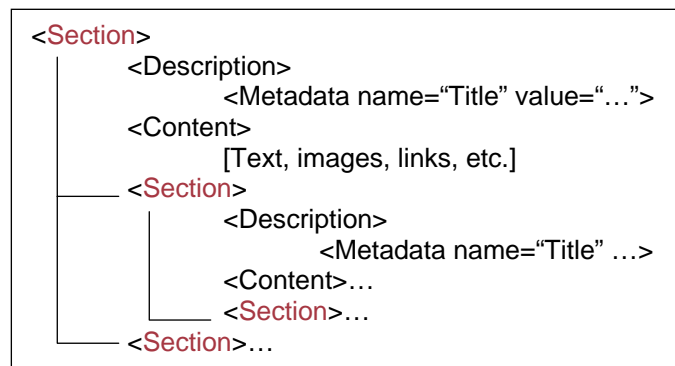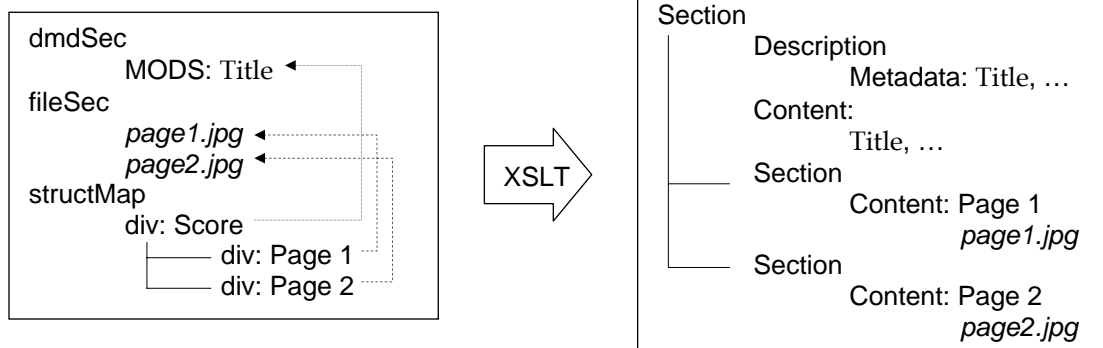
GSAF

Documents in Greestone 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 searched at the section level.

```
<Section>
        <Description>
                <Metadata name="Title" value="…">
        <Content>
                [Text, images, links, etc.]
        <Section>
                <Description>
                        <Metadata name="Title" …>
                <Content>…
                <Section>…
        <Section>…
```

METS to GSAF

```
dmdSec                                  Section
        MODS: Title                             Description
fileSec                                                 Metadata: Title, …
        page1.jpg                              Content:
        page2.jpg                                      Title, …
structMap                                      Section
        div: Score                                     Content: Page 1
                div: Page 1                                    page1.jpg
                div: Page 2                            Section
                                                               Content: Page 2
                        XSLT                                   page2.jpg
```

The top level div represents the score as a whole, its children represent the pages. The MODS record for the score is transformed into the metadata for searching and browsing the collection, and is inserted into the top-level Description; 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 in the GSAF file.

Links & Further Information:

Chopin Early Editions: http://chopin.lib.uchicago.edu/

Greenstone: http://www.greenstone.org/

      Downloads, documentation, examples; Photograph Album, Lehigh Digital Bridges, and
      WRLC collections shown in the presentation may be found here.

New Zealand Digital Library Project: http://www.nzdl.org/

      UNESCO & related collections, many demos; Virtual Disaster Library may be found here.

Witten, Ian H. & Bainbridge, David. *How to Build a Digital Library*. Morgan Kaufman, 2003.

Materials from this talk: http://www.lib.uchicago.edu/dldc/talks/2003/dlf-greenstone/