METS PROFILE

|  |  |
| --- | --- |
| Title: | **Boston College Text Objects** |
| Abstract: | This is a profile for METS objects that represent textual material issued in a paged format. Textual material is defined as material that is *mostly* textual in nature; it may include images, maps, tables or other formats. |
| Revision Dates: | 8/2010; 6/2011; 2/2014 |
| Contact: | Betsy Post  Boston College  O’Neill Library  140 Commonwealth Ave.  Chestnut Hill, MA 02467  617-552-1989  betsy.post@bc.edu |
| Related Profile: |  |
| Extension Schema: | |  |  |  |  |  | | --- | --- | --- | --- | --- | | **MODS**   |  |  | | --- | --- | | Context | Used for encoding descriptive metadata | | Note | When a bib record for the print version exists in Aleph, export in MARC, then run through PERL script similar to internet archive script to generate record for e-version. Use BC custom MARCedit transformation to convert to MODS. Run step2.xslt. Edit access condition, internetMediaType, and digitalOrigin. <mods:accessCondition type="useAndReproduction">Use of this public-domain resource is unrestricted.</mods:accessCondition> Need to simplify this process. | | | **ALTO**   |  |  | | --- | --- | | Context | Usually used for OCR | | Note |  | | | **BCdigiprov**   |  |  | | --- | --- | | Context | A local xml schema used in amdSec\digiprovMD to describe the digital provenance of each fileGrp in the fileSec | | Note |  | | |
| Description Rules: | All applications of the MODS schema in conforming METS documents follow the MODS User Guidelines published by Library of Congress' Network Development and MARC Standards Office. |
| Controlled Vocabularies: | |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Boston College Digitized Text Objects <mets> TYPE attribute values**   |  |  | | --- | --- | | Maintenance Agency | Boston College | | Values | text-monograph-whole *(may be used for bound manuscripts)*  text-monograph-part  text-serial-journal-whole | | Context | mets/@TYPE | | Description | A hierarchical value based [Oxford Digital Library conventions](http://www2.odl.ox.ac.uk/guidelines/odl-types.html). Dashes rather than dots are used to separate hierarchies to accommodate Digitool i18n message properties values. | | | **Boston College Digitized Text Objects <fileGrp> USE attribute values**   |  |  | | --- | --- | | Maintenance Agency | Boston College | | Values | |  |  |  |  | | --- | --- | --- | --- | | archive image | reference image | alto | permissions *(ailleurs eg.)* | |  |  | pdf |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | | | Context | mets/fileSec/fileGrp/@USE | | Description | File group use attributes describe two things about the file group: (1) is the image intended for access and (2) the kind of content the file represents. A mapping table in digitool uses these values to create an object’s use value and entity type value. | | | **Boston College Digitized Text Objects <structMap> TYPE attribute values**   |  |  | | --- | --- | | Maintenance Agency | Boston College | | Values | * physical * logical * mixed * pdf [yes, lowercase] | | Context | mets/structMap/@TYPE | | Description | Purely logical structMaps will be rare because of the amount of data/effort required to create them. | | | **Boston College Digitized Text Objects <structMap> LABEL attribute values**   |  |  | | --- | --- | | Maintenance Agency | Boston College | | Values | * Read Online * PDF Version * PDF Version (Compressed) | | Context | mets/structMap/@LABEL | | Description | *Read Online* is used for when the content is delivered at the page level via a page turning device. ***PDF Version*** is used for when the object is represented by one or more PDFs that contain several pages each. When PDFs are created at the page level for an object that has more than one page, then *Read Online* is used. | | | **Boston College Digitized Text Objects <div> TYPE attribute values**   |  |  | | --- | --- | | Maintenance Agency | Boston College | | Values | |  | | --- | | **Content Model 1: Bound Monographic Material** | | Use: Use this model for collections of wood or ivory tables, or sheets of paper, parchment, or similar material, that are blank, written on, or printed, and are strung or bound together; commonly many folded and bound sheets containing continuous printing or writing (AAT book definition).   |  |  |  | | --- | --- | --- | | Physical | Logical | Mixed | | volumes | logical | volumes | | volume |  | volume | | page |  | page | |  |  | logical | | | | Context | structMap/div/@TYPE | | Description | The div/@TYPE values are based on the content model designed by the National Library of Australia. The *content* attribute can be used at any level of the hierarchy except for the first level. The remainder of the values should be used hierarchically. Different subsets or values are appropriate depending on the structMap/@TYPE | | |  | |
| Structural Requirements | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **metsRootElement**   |  |  | | --- | --- | | 1 | The root <mets> element must include an OBJID attribute value containing an identifier that uniquely identifies the object in its owning repository. | | 2 | The root <mets> element include a LABEL attribute value that is the title of the of the object/entity being described in the METS document for the user. The title should be formatted using the Chicago Manual of Style’s [Notes and Bibliography](http://www.chicagomanualofstyle.org/ch16/ch16_sec003.html) conventions. | | 3 | The root <mets> element should contain a PROFILE attribute indicating that it conforms to this profile (not implemented at this time) | | | **metsHdr** | | |  |  | | --- | --- | | 1 | The <metsHdr> element must include the CREATEDATE attribute value. It must also include the LASTMODDATE attribute value if this does not coincide with the CREATEDATE | | 2 | Conforming METS documents must contain a metsHdr element. | | 3 | The <metsHdr> element must include a child <agent> element identifying the person or institution responsible for creating the METS object. The <agent> element must include ROLE (or OTHERROLE) and TYPE (or OTHERTYPE) attributes. | |  | altRecordId??? For links to parent or appropriate to structLink | | | **dmdSec** | | |  |  | | --- | --- | | 1 | Each <dmdSec> must include an ID attribute | | 2 | Conforming METS documents may, but need not, contain a one or more <dmdSec> elements. Each <dmdSec> may in turn contain a <dmdRef> or a <dmdWrap> | | 3 | If a <dmdSec> of a conforming document contains a <dmdWrap> with <xmlData>, the <xmlData> must conform to the MODS schema. | | | **amdSec** | | |  |  | | --- | --- | | 1 | Conforming METS documents may but need not contain an <amdSec> element. This <amdSec> may but need not contain one or more <techMD> elements, <sourceMD> elements, <rightsMD> elements and/or <digiprovMD> elements. | | 2 | A conforming METS document will contain no more than one <amdSec> element. All <techMD>, <sourceMD>, <rightsMD> and <digiprovMD> elements will appear in this single <amdSec> element. | | 3 | If one or more <techMD> elements pertaining to image content files are present, they must contain <xmlData> of NISOIMG type conforming to the MIX schema. | | 4 | A digital copy of signed permission forms will be stored as base 64 bindata inside an <mdWrap>; the mdWrap must have an ID attribute | | 5 | A <digiprovMD> may be created for each file group to record information about the migration/transformation of the files in the group subsequent to the original digitization to its current incarnation as a digital object. | | 6 | Each <digiprovMD> element must include an ID attribute. | | 7 | Digital provenance md include a local BCprovenance element wrapped in <mdWrap><xmlData> | | | **fileSec**   |  |  | | --- | --- | | 1 | The <fileSec> of a conforming METS document must contain a <fileGrp> for each file format/use represented by the content files. For example, the <fileSec> of a typical METS document implementing this profile might contain one <fileGrp> representing TIFF master images, one <fileGrp> representing edited master TIFF images, and one <fileGrp> representing JPEG reference images. <fileGrp> elements may not contain subsidiary <fileGrp> elements. | | 2 | Each <fileGrp> represented in the <fileSec> must have an associated USE attribute. Supported <file>/<fileGrp> USE attribute values appear in the <controlled\_vocabularies> section of this document. | | 3 | Constituents of each <fileGrp> will represent equivalent units of content. For example, chapter level pdfs and a pdf for the whole work would be represented in separate file groups. | | 4 | Each <fileGrp> will have a VERSDATE for the file grouping | | 5 | Each <file> may have an associated ADMDID attribute to identify the associated <digiprovMD> | | 6 | Each file must have the following attributes: ID, MIMETYPE, GROUPID, and SEQUENCE | | 7 | Each file may have the following attributes CHECKSUM, CHECKSUMTYPE, SIZE | |  |  | | | **structMap** | | |  |  | | --- | --- | | 1 | A conforming METS document may contain multiple <structMap> elements. | | 2 | Constituent files in a <structMap> will represent equivalent units of content. For example, a page image representation and a pdf of an entire book would be represented in separate structMaps. | | 3 | A conforming <structMap> will include an ID attribute. | | 4 | A conforming <structMap> will include LABEL and TYPE attributes using terms from the controlled vocabularies section of this profile. | | 5 | Each <div> that corresponds to a descriptive metadata record will have a DMDID attribute that refers to the record. | | 6 | Each <div> will have a TYPE attribute as defined in the controlled vocabularies section of this document. | | 7 | Each <div> must include a LABEL attribute value. | | 8 | Each <div> will contain an ID attribute. (Note: The ID attribute will be generated by Digitool at the time of ingest and any existing ID attributes will be overwritten upon ingest. This makes referring to specific <div> elements in the structMap difficult, so a workaround is used in the <behaviorSec>) | | | **behaviorSec** | | |  |  | | --- | --- | | 1 | The <behaviorSec> STRUCTID attribute will identify the <file> ID for the reference image which should be the default opening page. (The STRUCTID should actually contain the relevant <div> ID attributes for each structMap. However, this is not practical in the Digitool context as div\@ID attributes are overwritten on ingest.) | |  |  | |  |  | | | **applicationStuff** | | When a page turning application that shows facing pages encounters a structMap/div with the TYPE volume, it will know that the first page image should be displayed on its own and that subsequent pairs of pages should be displayed facing eachother. | | **Digitool Ingest Instructions** | | The task chain should include (1) technical metadata extraction and (2) control section attribute assignment. | | In the parameters set, the value for any set of files that couldn’t be easily regenerated be a batch process should have the preservation\_level set to *critical*. Critical preservation items would include masters, intermediate masters, alto, mets, etc. | | On ingest, the mets\@TYPE attribute is prefixed to the digital entity label for the METS xml object. (Ideally, I think the system would take this value from MODS:genre.) | | Update the entity’s USE value to VIEW\_MAIN. | | Create a thumbnail right away. | |