

DRS Content Guide to In-Production Content Models

Last modified: 01/26/18

Table of Contents

Part 1: Guide to this document.....	2
Changes to content model definitions	2
Content model definition elements	2
Part 2: Introduction to DRS objects	4
Data and content models	4
Technical restrictions.....	4
Controlled vocabulary	4
Object descriptors.....	5
Part 3: Content model definitions	7
CMID-1.0 (OPAQUE) Status - IN-PRODUCTION	7
CMID-1.1 (OPAQUE CONTAINER) Status - IN-PRODUCTION	10
CMID-2.0 (AUDIO) Status - IN-PRODUCTION	14
CMID-3.0 (WEB HARVEST) Status - IN-PRODUCTION	22
CMID-4.0 (PDS DOCUMENT) Status - IN-PRODUCTION	25
CMID-4.1 (DOCUMENT) Status - IN-PRODUCTION	30
CMID-4.3 (GOOGLE DOCUMENT CONTAINER 1) Status – IN-PRODUCTION	35
CMID-4.4 (GOOGLE DOCUMENT CONTAINER 2) Status – IN-PRODUCTION	38
CMID-4.6 (GOOGLE DOCUMENT CONTAINER 3) Status – IN-PRODUCTION	41
CMID-5.0 (STILL IMAGE) Status - IN-PRODUCTION	44
CMID-5.2 (TARGET IMAGE) Status - IN-PRODUCTION	48
CMID-6.0 (TEXT) Status - IN-PRODUCTION.....	51
CMID-8.0 (COLOR PROFILE) Status - IN-PRODUCTION	54
CMID-9.0 (EMAIL MESSAGE) Status - IN-PRODUCTION.....	57
CMID-10.0 (COLLECTION) Status - IN-PRODUCTION	60
CMID-11.0 (PACKET) Status - IN-PRODUCTION.....	61
CMID-12.0 (VIDEO) Status - IN-PRODUCTION.....	63
CMID-14.0 (PDS DOCUMENT LIST) Status - IN-PRODUCTION.....	70

Part 1: Guide to this document

Changes to content model definitions

All changes made to completed content model definitions¹ must be backwards compatible. In cases where non-backwards compatible changes must be made, a new version of the content model must be defined. This requirement ensures that modifications to content model definitions do not cause objects to become invalid according to the object's content model definition.

Content model definition elements

Each Content Model Definition has the following elements:

- **Definition status/Completion status** = status of the completeness of the definition, one of:
 - DRAFT = The definition is not yet complete.
 - COMPLETED = The definition is complete.
- **Definition status/Ingest support status** = status of the production version of the DRS ingest support of the definition, one of:
 - IN-DEVELOPMENT = The content model is in design stage or is only supported by the Dev version of the DRS.
 - IN-QA = The content model is only supported by the QA version of the DRS.
 - IN-PRODUCTION = The content model is supported by the production version of the DRS. **Only these content models are contained in this document.**
 - DEPRECATED = The content model is scheduled to become unsupported. The production version of the DRS will still accept objects under this content model but a date has been set to withdraw this support.
 - UNSUPPORTED = The content model is no longer supported for new ingest by the production version of the DRS.
- **Identity/Content model ID** = an identifier that uniquely identifies the content model
 - The format of the ID is CMID-[x].[y] where x is a number that identifies the content model and y identifies the version of the content model. For example, CMID-3.1, CMID-3.2 and CMID-3.3 identify three similar content models. CMID-3.3 is a new version of CMID-3.2, and CMID-3.2 is a new version of CMID 3.1. New versions are created when non-backwards compatible changes must be made to the definition.
- **Identity/Aliases** = One or more descriptive names for the content models
- **Summary description/Description** = a brief description of the content model
- **Summary description/Compatible object genres** = One or more categories of digital objects that are well modeled by the content model
- **Related software and systems/Delivery applications** = All HL-managed applications that can

¹ Completed content model definitions have a completion status of COMPLETED.

accept delivery requests for objects, or representations of the object, from users and deliver them to the requesting user.

- The DRS Web Admin is counted as a delivery application in this document but only supports the real-time download of a single file or object, or set of files or objects at a time and requires the user to first perform a search and then select what to download.
- **Related software and systems/Rendering applications** = A selected set of known applications that can render one or more representations of objects in this content model; where possible, applications are selected from each of the following areas: (1) popular software users are likely to have access to, and (2) open source software. The applications should account for all popular platforms (devices and operating systems).
- **Assessments and preservation plans** = In-line or references to assessments or other preservation plans for objects conforming to the content model
- **Content model details/Structure** = A diagram showing the arrangement, obligation, cardinality, and internal-to-the-object relationships for objects that conform to this content model
 - It is assumed in these diagrams that the object's content files have not been deleted. Note that it is possible for a curator to delete an object in which case the object will contain an object descriptor file and no other files.
- **Content model details/Relationships** = Detailed information about the kinds of relationships that can exist to or from files and objects in this content model
- **Content model details/Relationship diagram** = A diagram showing all the potential internal and external relationships from or to the files or the object
- **Content model details/Roles** = The object- and file- level roles that are valid for this content model
- **Content model details/Directory structure** = Information about what is known about the directory structure of objects that conform to this content model. Directory structures are not prescribed by the DRS but there tend to be consistencies because of depositor tools and unchanging workflows, and HL-provided deposit tools (Batch Builder 1 and 2, DMART).
- **Content model details/Persistent names** = How HL tools handle the assignment of URNs for files and objects in this content model
- **Content model details/Metadata requirements** = Any format, technical or other metadata requirements for files within objects that conform to this content model
- **Content model details/Descriptor details** = Information that is specific to object descriptors for this content model. Information about descriptors that apply to all content models is found in Part 2 of this document.
- **Open questions** = While a content model definition is in draft or being conserved for revision, this section will contain the questions that need to be resolved.
- **Change history** = A record of all changes made to the definition, starting from the definition completion date (when completion status became equal to COMPLETED). If the definition has been deprecated, the date at which the definition's ingest support status will become unsupported should be included here.

Part 2: Introduction to DRS objects

Data and content models

DRS content is modeled at three levels of granularity: object, file and bitstream.

- An object in the DRS is a coherent set of content that is considered a single intellectual unit for purposes of description, use and/or management: for example a particular book, web harvest, serial or photograph.
- A file is a named and ordered sequence of bytes that is known to an operating system. A file can have zero or more bytes and has a file format, access permissions, and file system characterizations such as file size and last modification date.
- A bitstream is contiguous or non-contiguous data within a file that has meaningful properties for preservation or access purposes.

DRS objects are composed of one or more files. Files are composed of zero or more bitstreams. A file can be contained by at most a single object. Files are not shared among objects - each file can belong to a single object. A bitstream may contain other bitstreams but can be contained by at most a single file. While DRS material is always modeled at the object and file levels, material is only modeled at the bitstream level when it is deemed necessary for preservation or access purposes.

All objects in the DRS conform to a single content model. Content models define types of objects supported by the DRS. Each content model definition documents:

- supported file formats
- file and object relationships
- file roles and other metadata
- known delivery and rendering applications
- associated assessments and preservation plans

Technical restrictions

The following restrictions applied to all DRS content in the old Web Admin:

- Maximum file size: tested up to 20 GB for deposit and up to 50 GB for download from WebAdmin

It's not clear yet what the limits for download are in the new Web Admin.

Controlled vocabulary

There are several locations within DRS systems, schemas and code where enumerated lists of controlled

vocabulary are maintained:

- WordShack
- XML schemas, particularly HulDrsAdmin
- An OTS class: Descriptor
 - For the source code see <http://viewvc.hul.harvard.edu:1234/viewvc/Development/OTS/src/edu/harvard/hul/ois/ots/Descriptor.java?view=markup> (access is restricted to repository staff)

Object descriptors

Each object has a single metadata file called an object descriptor that contains descriptive, administrative, preservation, technical and structural metadata about the object, its files and bitstreams. This descriptor is formatted in METS XML format. The file is given the metadata role OBJECT_DESCRIPTOR.

These descriptor details are common to all content models:

- mets element
 - TYPE attribute indicates the content model, one of:
 - OPAQUE
 - OPAQUE CONTAINER
 - AUDIO
 - WEB HARVEST
 - PDS DOCUMENT
 - DOCUMENT
 - GOOGLE DOCUMENT CONTAINER 1
 - GOOGLE DOCUMENT CONTAINER 2
 - GOOGLE DOCUMENT CONTAINER 3
 - STILL IMAGE
 - BIOMEDICAL IMAGE
 - TARGET IMAGE
 - TEXT
 - COLOR PROFILE
 - EMAIL MESSAGE
 - COLLECTION
 - PACKET
 - VIDEO
 - PDS DOCUMENT LIST
 - PROFILE attribute = “HUL”
 - OBJID attribute contains the persistent identifier for the object
- metsHdr element
 - agent element
 - name element has value “Harvard University Library”
- may have a dmdSec element containing mods metadata
- a series of amdSec sections, not necessarily in this order:
 - an amdSec/techMd section with object identifiers in the premis:object schema
 - one objectIdentifierType with value HUL_DRS_OBJECT_ORACLE

- This is the Oracle system ID for the object
 - one objectIdentifierType with value HUL_DRS_DESCRIPTOR_ORACLE
 - This is the Oracle system ID for the object descriptor file
 - one objectIdentifierType with value HUL_DRS_OBJECT_URN
 - This is the persistent identifier for the object
 - one originalName containing the object's owner supplied name
 - an amdSec/techMd section containing administrative metadata for the object in the hulDrsAdmin schema
 - amdSec/techMd section containing administrative metadata for each file in the hulDrsAdmin schema
 - an amdSec/techMD section per file containing preservation metadata in the premis:schema schema
 - May have format-specific technical metadata in the premis:objectCharacteristicsExtension element
 - May have any number of premis:relationship elements documenting relationships to or from the file
 - an amdSec/digiprovMD section containing events in the premis:event schema
- fileSec section containing a listing of all files belonging to the object, grouped by MIMIE-type, for example:


```

<fileGrp ID="ID109" USE="application-msword">
  <file ID="ID116" ADMID="ID113 ID110" MIMETYPE="application/msword">
    <FLocat ID="ID117" LOCTYPE="OTHER" OTHERLOCTYPE="HUL_DRS_FILE_ORACLE"
xlink:href="2022937"/>
  </file>
  <file ID="ID133" ADMID="ID130 ID127" MIMETYPE="application/msword">
    <FLocat ID="ID134" LOCTYPE="OTHER" OTHERLOCTYPE="HUL_DRS_FILE_ORACLE"
xlink:href="2022942"/>
  </file>
</fileGrp>

```
- Unless indicated by the content model definition, structmap@TYPE and structmap/div@TYPE are not used.

Part 3: Content model definitions

CMID-1.0 (OPAQUE) *Status - IN-PRODUCTION*

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** OPAQUE
- **Content Model ID:** CMID-1.0

Summary Description

- **Description:** This content model represents an object containing files in any format. There are no intellectual restrictions on what the object can represent. For instance, it could represent a single intellectual work (e.g. a video game), relate to an individual (e.g. all donated works from a faculty member) or be associated with a particular medium (e.g. selected content from a hard drive). The technical restrictions that apply to all DRS objects also apply to opaque objects (See the [technical restrictions section](#) in this document).
- **Compatible object genres:** Any object genre not currently supported by the DRS (e.g. video); any objects formatted differently than expected by HL software (e.g. websites collected with different tools than WAX uses); large sets of unprocessed files (e.g. content of a hard drive)

Related Software and Systems

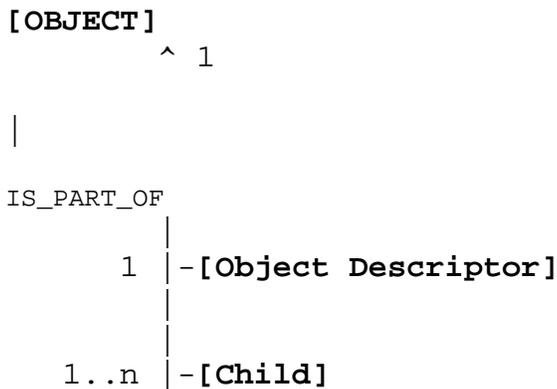
- **Delivery applications:** DRS Web Admin, FDS
- **Rendering applications:** none

Assessments and Preservation Plans

- No current assessments or plans

CONTENT MODEL DETAILS

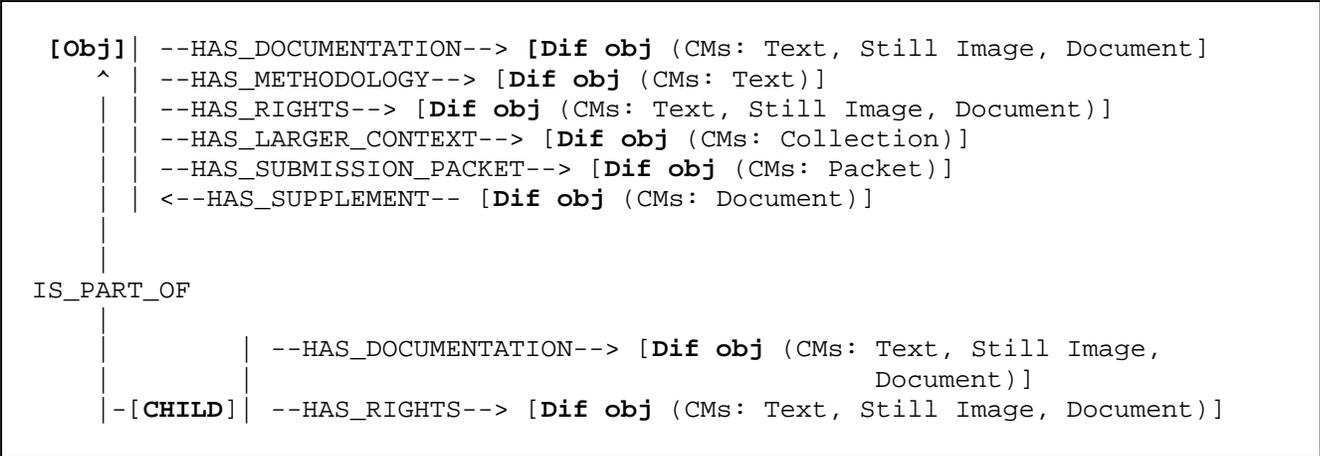
Structure: This object is composed of one or more child files in any format. The child files are one of two types: content files or documentation files. The content files are the focus of preservation; the documentation files contain information about the content files.



Structure diagram for an opaque object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

Relationships (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)

- File-to-its Object
 - IS_PART_OF relationships between each child and the parent
 - These are implicitly modeled by describing the files in the object descriptor.
- File-to-File within the same Object
 - No relationships.
- File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object (as in the case of email attachment objects)
 - May be the target object of a HAS_SUPPLEMENT relationship from another object.
 - May be the source object of a HAS_SUBMISSION_PACKET relationship between this attachment object and its associated packet object.
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in an opaque object, or the opaque object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

Roles

- Object (may have one or more of the following roles)
 - EMAIL_ATTACHMENT
 - THESIS_SUPPLEMENT
- Files (may have one or more of the following roles)
 - DOCUMENTATION
 - LICENSE
 - ORIGINAL_ORDER

Directory structure

- All the content files are contained in one of two top-level directories:
 - content
 - This directory contains all content files except DRS documentation files.
 - documentation
 - This directory contains all DRS documentation files (original order, license, etc.)

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: no URNs requested for the content files

Metadata Requirements

Child (content file)	<ul style="list-style-type: none"> ● Acceptable Formats <ul style="list-style-type: none"> ○ Any format <ul style="list-style-type: none"> ■ format name for unidentified formats: Unknown Binary ○ MIME type: any <ul style="list-style-type: none"> ■ MIME type for unidentified formats: application/octet-stream ● Format technical metadata <ul style="list-style-type: none"> ○ The type of technical metadata (text, audio, etc.) is determined by the DRS metadata extraction tool, FITS, based on a mapping of format names to metadata types.
----------------------	--

	<ul style="list-style-type: none"> ● File metadata <ul style="list-style-type: none"> ○ Usage class: LOWUSE (suggested default) ○ Access flag: N (suggested default)
Child (documentation file)	<ul style="list-style-type: none"> ● Format: any of the following: <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ▪ Format name: Plain Text ▪ MIME type: text/plain ▪ Character encoding: US-ASCII or UTF-8 ▪ Technical metadata type: TEXT ○ 2 <ul style="list-style-type: none"> ▪ Format name: OpenDocument Text ▪ MIME type: application/vnd.oasis.opendocument.text ▪ Technical metadata type: DOCUMENT ○ 3 <ul style="list-style-type: none"> ▪ Format name: PDF/A-1a <ul style="list-style-type: none"> ● although this is preferred, Portable Document Format or any of the other PDF/A variations is acceptable ▪ MIME type: application/pdf ▪ Technical metadata type: DOCUMENT ○ 4 <ul style="list-style-type: none"> ▪ Format name: OpenDocument Spreadsheet ▪ MIME type: application/vnd.oasis.opendocument.spreadsheet ▪ Technical metadata type: N/A

Descriptor Details

- mets element
 - TYPE attribute = "OPAQUE"
- fileSec
 - FileGrp by MIME-type, for example:

```
<fileGrp ID="ID109" USE="application-msword">
  <file ID="ID116" ADMID="ID113 ID110" MIMETYPE="application/msword">
    <FLocat ID="ID117" LOCTYPE="OTHER" OTHERLOCTYPE="HUL_DRS_FILE_ORACLE" xlink:href="2022937"/>
  </file>
  <file ID="ID133" ADMID="ID130 ID127" MIMETYPE="application/msword">
    <FLocat ID="ID134" LOCTYPE="OTHER" OTHERLOCTYPE="HUL_DRS_FILE_ORACLE" xlink:href="2022942"/>
  </file>
</fileGrp>
```
- structMap contains div elements by file type (content file or documentation file)
 - At most one div with TYPE="CONTENT"
 - At most one div element with TYPE="DOCUMENTATION"

CMID-1.1 (OPAQUE CONTAINER) Status - IN-PRODUCTION

Definition Status

- Completion status: COMPLETED
- Ingest support status: IN-PRODUCTION

Identity

- **Aliases:** OPAQUE CONTAINER
- **Content Model ID:** CMID-1.1

Summary Description

- **Description:** This content model represents an Opaque (CMID-1.0) object if the object were unzipped. The object's files are archived into a single compressed Zip file. Within the Zip file there can be files in any format. There are no intellectual restrictions on what the object can represent. For instance, it could represent a single intellectual work (e.g. a video game), relate to an individual (e.g. all donated works from a faculty member) or be associated with a particular medium (e.g. selected content from a hard drive). The technical restrictions that apply to all DRS objects also apply to opaque objects (See the [technical restrictions section](#) in this document).
- **Compatible object genres:** Any object genre not currently supported by the DRS (e.g. software); any objects formatted differently than expected by HL software (e.g. websites collected with different tools than WAX uses); large sets of unprocessed files (e.g. content of a hard drive)

Related Software and Systems

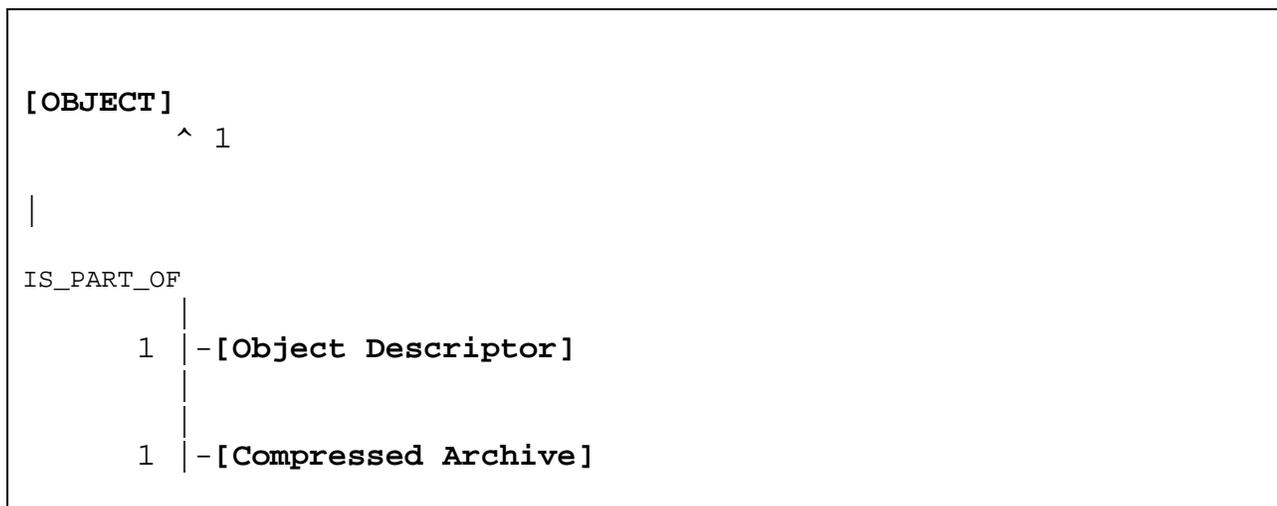
- **Delivery applications:** DRS Web Admin, FDS
- **Rendering applications:** none

Assessment and Preservation Plans

- No current assessments or plans

CONTENT MODEL DETAILS

Structure: This object is composed of a single content file: a compressed archive containing content files and optionally documentation.



Structure diagram for an opaque container object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object

- This directory contains content files (not documentation about the content).
- documentation
 - This directory contains DRS documentation files (original order, license, etc.)

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: no URNs requested for the compressed archive file.

Metadata Requirements

Compressed Archive	<ul style="list-style-type: none"> ● Acceptable Formats <ul style="list-style-type: none"> ○ Format name: ZIP ○ MIME type: application/zip ● Format technical metadata: containerMD ● File metadata <ul style="list-style-type: none"> ○ Usage class: LOWUSE (suggested default) ○ Access flag: N (suggested default)
Content files within the compressed archive (bitstreams)	<ul style="list-style-type: none"> ● Allowable Formats <ul style="list-style-type: none"> ○ Any format <ul style="list-style-type: none"> ■ format name for unidentified formats: Unknown Binary ○ MIME type: any <ul style="list-style-type: none"> ■ MIME type for unidentified formats: application/octet-stream ● Format technical metadata: none ● File metadata: N/A (these are bitstreams)
DRS documentation files within the compressed archive (bitstreams)	<ul style="list-style-type: none"> ● Format: any of the following: <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ■ Format name: Plain Text ■ MIME type: text/plain ■ Character encoding: (not recorded for bitstreams) ■ Technical metadata type: none ○ 2 <ul style="list-style-type: none"> ■ Format name: OpenDocument Text ■ MIME type: application/vnd.oasis.opendocument.text ■ Technical metadata type: none ○ 3 <ul style="list-style-type: none"> ■ Format name: PDF/A-1a <ul style="list-style-type: none"> ● although this is preferred, Portable Document Format or any of the other PDF/A variations are acceptable ■ MIME type: application/pdf ■ Technical metadata type: none ○ 4 <ul style="list-style-type: none"> ■ Format name: OpenDocument Spreadsheet ■ MIME type: application/vnd.oasis.opendocument.spreadsheet ■ Technical metadata type: none

Descriptor Details

- mets element

- TYPE attribute = “OPAQUE CONTAINER”
- fileSec
 - FileGrp by MIME-type, for example:


```
<fileGrp ID="ID109" USE="application-zip">
  <file ID="ID116" ADMID="ID113 ID110" MIMETYPE="application/msword">
    <FLocat ID="ID117" LOCTYPE="OTHER" OTHERLOCTYPE="HUL_DRS_FILE_ORACLE" xlink:href="2022937"/>
  </file>
</fileGrp>
```

CMID-2.0 (AUDIO) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** AUDIO
- **Content Model ID:** CMID-2.0

Summary Description

- **Description:** This content model represents sound works in digital form. It may include auxiliary content including process history files.
- **Compatible object genres:** interview, speech, reading, musical work, conversation, lecture, drama, email attachments

Related Software and Systems

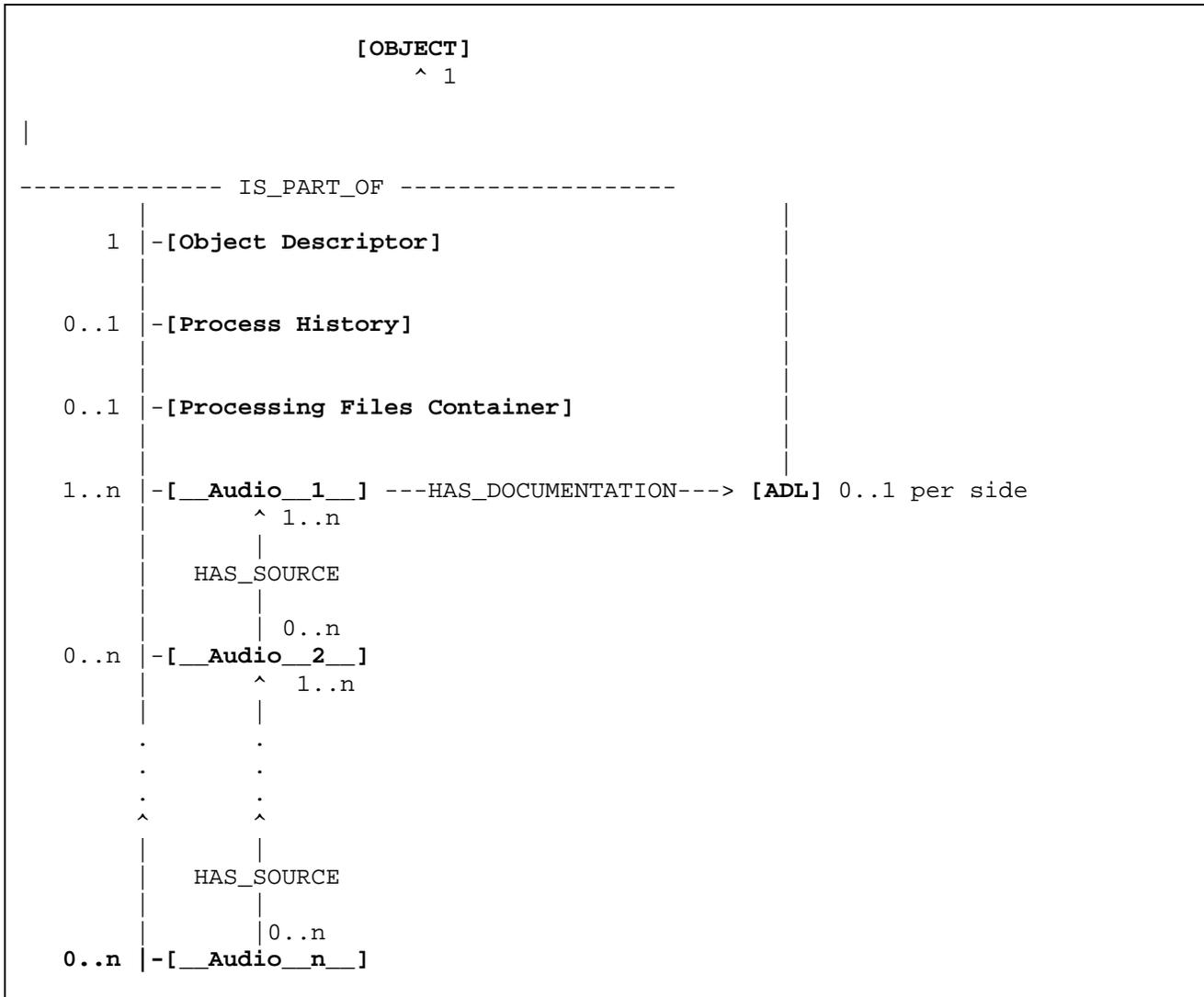
- **Delivery applications:** DRS Web Admin, SDS using the Wowza streaming server (MP3, MP4/AAC formats only), SDS using Helix (RealAudio format)
- **Rendering applications:** JWPlayer in a web browser (MP3, MP4/AAC formats only); RealPlayer for RealAudio files

Assessments and Preservation Plans

- Upcoming plan to migrate the RealAudio files to a different delivery format – MP3.

Content Model Details

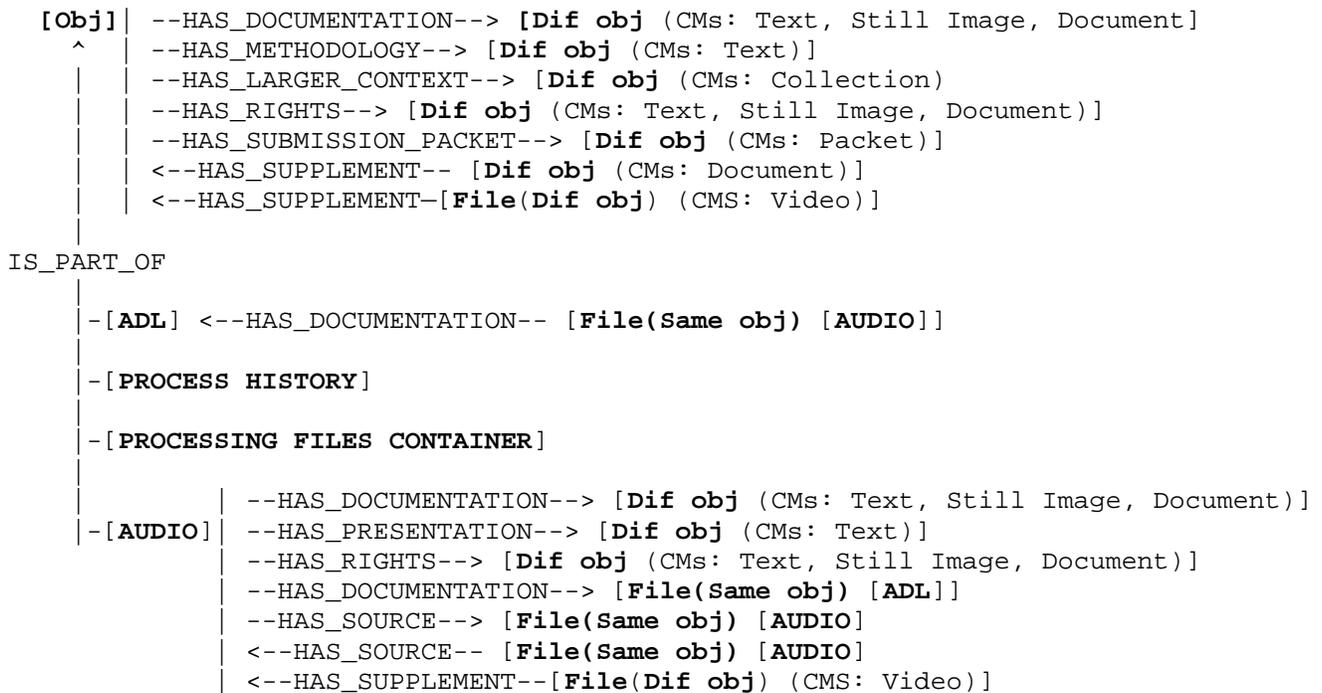
- **Structure:** One or more audio files related by derivative relationships, with optional related audio decision lists and process history files.



Structure diagram for an audio object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - Key relationships: Some of the audio files (especially those with ARCHIVAL_MASTER or PRODUCTION_MASTER roles) may be associated with ADL (Audio Decision List) files. The deliverable audio files may be associated with presentation objects (containing AES60 files) to provide delivery sequencing.
 - File-to-its Object
 - IS_PART_OF relationships between each child and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_SOURCE relationships between derivative audio files. **Note that multiple audio files can be the source of another audio file, for example when multiple archival masters were combined to form a production master.**
 - HAS_DOCUMENTATION relationships between audio content files and ADL files.

- File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - The deliverable files may be the source of HAS_PRESENTATION relationships to AES60 file objects.
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object (as in the case of email attachment objects)
 - May be the target object of a HAS_SUPPLEMENT relationship from another object.
 - May be the source object of a HAS_SUBMISSION_PACKET relationship between this attachment object and its associated packet object.
- Object-to-File in Different object
 - The object may be the target of HAS_SUPPLEMENT relationships from video files in the video content model.
- File-to-File within Different Object
 - The content files may be the target of HAS_SUPPLEMENT relationships from video files in the video content model.



Relationship diagram showing what is possible in terms of relationships between files in an audio object, or the audio object itself, and other files or objects. (Key: Dif obj = Different object; File (Same obj) = File within the same object; CMs = Content Models)

- **Roles**

- Object may have any of the following roles:

- EMAIL_ATTACHMENT
- THESIS_SUPPLEMENT

- Files

- The audio content files are differentiated by their role: ARCHIVAL_MASTER, PRODUCTION_MASTER or DELIVERABLE. Typically audio files will have only one of these roles, but it is possible for them to have multiple roles, as in the case of archival masters that are in a format that can be delivered.
- The ADL file has the role AUDIO_DECISION_LIST
- The Process History file has the role PROCESS_HISTORY
- The Processing Files Container has the role PROCESS_FILES

- **Directory structure**

- Directory structure will vary depending on the deposit tool used.

- DMART (The underlined files and directories must be named as shown.)

workId/

archival/

segmentId/

(archival master audio files)

production/

segmentId/

(production master audio files)

deliverable/

segmentId/

(deliverable audio files)

- Batch Builder 2 (simple audio)
 - Audio files are contained in directories that begin with the prefix audio
- Batch Builder 2 (advanced audio)
 - A sample directory structure is shown below. The underlined files and directories must be named as shown.

audio.properties (required)

relationships.csv (required if derivative relationships)

mods.xml (optional; descriptive metadata for the object)

my_object-A299ff/

archival/ (required if there are archival master files)

digiprov.xml (process history for object)

sub_dir1/

file1.wav

file1.xml (AES57 metadata for file1.wav)

file1.properties (DRS metadata for file1.wav)

file2.wav

file2.xml (AES57 metadata for file2.wav)

file2.properties (DRS metadata for file2.wav)

sub_dir1.adl (ADL for *.wav files in sub_dir1)

sub_dir2/

file1.wav

file1.xml (AES 57 metadata for file1.wav)

file1.properties (DRS metadata for file1.wav)

file2.wav

file2.xml (AES57 metadata for file2.wav)

file2.properties (DRS metadata for file2.wav)

sub_dir2.adl (ADL for *.wav files in sub_dir2)

deliverable/
performances/ (required if have deliverable files)

file1.mp3

file1.xml (AES57 metadata for file1.mp3)
file1.properties (DRS metadata for file1.mp3)
file2.mp3
file2.xml (AES57 metadata for file2.mp3)
file2.properties (DRS metadata for file2.mp3)
playlists/ (required for playlists)
playlist1.xml (AES60 file)
playlist2.xml (AES60 file)

misc/ (optional; contains any # of files in any format)

original/ (optional)

audio_object.xml (AES57 metadata for original item)

production/ (required if have production master files)

sub_dir1/
file1.wav
file1.xml (AES57 metadata for file1.wav)
file1.properties (DRS metadata for file1.wav)
file2.wav
file2.xml (AES57 metadata for file2.wav)
file2.properties (DRS metadata for file2.wav)
sub_dir1.adl (ADL for *.wav files in sub_dir1)
sub_dir2/
file1.wav
file1.xml (AES57 metadata for file1.wav)
file1.properties (DRS metadata for file1.wav)
file2.wav
file2.xml (AES57 metadata for file2.wav)
file2.properties (DRS metadata for file2.wav)
sub_dir2.adl (ADL for *.wav files in sub_dir2)

workspace/ (optional; can contain any # of files in any format)

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: request URNs only for the content files having the role DELIVERABLE

Metadata Requirements

Object component	Metadata requirements
Audio content files	<ul style="list-style-type: none">• Acceptable Formats:<ul style="list-style-type: none">○ 1<ul style="list-style-type: none">▪ Format name: Audio Interchange File Format▪ MIME type: audio/x-aiff○ 2<ul style="list-style-type: none">▪ Format name: Waveform Audio<ul style="list-style-type: none">• Technically these should be in Broadcast Waveform Audio format, although they are recognized by FITS as

	<p style="text-align: center;">Waveform Audio.</p> <ul style="list-style-type: none"> ▪ MIME type: audio/x-wave ○ 3 (won't support this format for ingest after the migration to DRS 2; it will exist in the legacy objects for some time after the migration from RealAudio to MP3) <ul style="list-style-type: none"> ▪ Format name: RealAudio ▪ MIME type: audio/vnd.rn-realaudio ○ 4 <ul style="list-style-type: none"> ▪ Format name: MPEG 1/2 Audio Layer 3 ▪ MIME type: audio/mpeg ○ 5 <ul style="list-style-type: none"> ▪ Format name: MPEG Advanced Audio Encoding ▪ MIME type: audio/mp4 ● File metadata <ul style="list-style-type: none"> ○ Role: one or more of: <ul style="list-style-type: none"> ▪ ARCHIVAL_MASTER <ul style="list-style-type: none"> ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ▪ PRODUCTION_MASTER <ul style="list-style-type: none"> ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ▪ DELIVERABLE <ul style="list-style-type: none"> ● Usage class: HIGHUSE (suggested default) ● Access flag: P (suggested default)
Process History file	<ul style="list-style-type: none"> ● Format: Extensible Markup Language <ul style="list-style-type: none"> ○ Process history is described by AES-x098C: Administrative metadata for audio objects - Process history schema (which hasn't been officially published yet). This file documents the complete history of the pre-DRS processing performed on this audio object, including the digitization process for the master files, and the devices and settings used. ● MIME type: text/xml ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: PROCESS_HISTORY
Processing Files Container	<ul style="list-style-type: none"> ● Format name: ZIP ● MIME type: application/zip ● Metadata type: none ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: PROCESS_FILES (suggested default) <p>This ZIP file contains information about intermediate audio content files that were created as part of the processing but were deleted prior to deposit of the audio content to the DRS. This intermediate file metadata is included in this</p>

	<p>audio object to provide information to the audio processing lab about their pre-DRS processing. It is not expected that the DRS will know very much about these files. This ZIP file contains one or more files in the following formats:</p> <ul style="list-style-type: none"> ● 1. <ul style="list-style-type: none"> ○ Format name: Extensible Markup Language ○ MIME type: text/xml ● 2. <ul style="list-style-type: none"> ○ Format name: Audio Decision List <ul style="list-style-type: none"> ■ ADL is described by AES31-3-2008: AES standard for network and file transfer of audio - Audio-file transfer and exchange - Part 3: Simple project interchange (Audio Decision List) ○ MIME type: text/plain? ● 3. <ul style="list-style-type: none"> ○ Format name: EDL? ○ MIME type: ? ● 4. <ul style="list-style-type: none"> ○ Format name: Unknown Binary ○ MIME type: application/octet-stream
Audio Decision List	<ul style="list-style-type: none"> ● Format name: Audio Decision List <ul style="list-style-type: none"> ○ ADL is described by AES31-3-2008: AES standard for network and file transfer of audio - Audio-file transfer and exchange - Part 3: Simple project interchange ○ According to the AES website: <ul style="list-style-type: none"> ■ “This standard provides a convention for expressing edit data in text form in a manner that enables simple and accurate computer parsing while retaining human readability. It also describes a method for expressing time-code information in character notation. It supports common professional audio sampling frequencies, video frame rates, and film framing. This document addresses the core need of the AES31 series of standards in providing a simple but extensible system for passing audio material between systems.” ○ ADL is an ASCII-text based format that uses the Edit Decision Markup Language (EDML), described in the same standard ○ There were two different versions deposited to the DRS but both were coded as version 1.00. The earlier version lacked the pan list and marker list. ● MIME type: text-x-adl ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: AUDIO_DECISION_LIST
Presentation	<ul style="list-style-type: none"> ● Format name: Extensible Markup Language ● MIME type: text/xml ● Usage class: HIGHUSE (suggested default) ● Access flag: P (suggested default)

Descriptor Details²

- mets element
 - TYPE attribute = “AUDIO”
- In addition to the amdSec sections common to all content files:
 - Audio content files will have audio technical metadata in the premis:objectCharacteristicsExtension element
 - at most one amdSec/sourceMD section containing audio object metadata for the analog source original
 - an amdSec/techMD section per audio decision list
 - this has an mdRef@OTHERMDTYPE = “AES-31”
 - at most a single amdSec/digiprovMD section for the audio process history file (digiprov.xml)
 - this has an mdRef@OTHERMDTYPE = “AUDIO_PROCESSING”
 - digiprovMD section with Premis events
- fileSec section (as specified for all content models)
 - This differs from the DRS1 descriptors where the ID attribute of the fileGrp element contains values such as Real, Wave, AES31, AIFF, Misc
 - What was included in the DRS1 descriptor Misc fileGrp will now be in the zipped Audio Process Files container (audio object metadata for files deleted prior to DRS ingest, ADLs for files deleted prior to DRS ingest, the precursors to ADLs for files deleted prior to ingest).
- structMap section
 - TYPE attribute = “PHYSICAL”³
 - Div sections with attribute TYPE⁴ one of” “ARCHIVAL”, “PRODUCTION” or “DELIVERABLE”

CMID-3.0 (WEB HARVEST) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** WEB HARVEST (preferred), WAX 1.0 CONTAINER
- **Content Model ID:** CMID-3.0

Summary Description

- **Description:** This content model represents the web harvest object deposited into the DRS by the HL Web Archiving Collection Service (WAX). The objects are prepared and deposited by the WAX Archiver using the Batch Builder command line interface.
- **Compatible object genres:** web harvest

²DRS will now omit techMD sections for audio creator metadata (i.e. Mac creator codes) that was present in earlier DRS audio models, and which used <http://hul.harvard.edu/ois/xml/xsd/drs/audioCreator.xsd>.

³ This is a change from DRS1 audio descriptors that use LOGICAL

⁴ This is a change from DRS1 audio descriptors that use the attribute LABEL for these values

Related Software and Systems

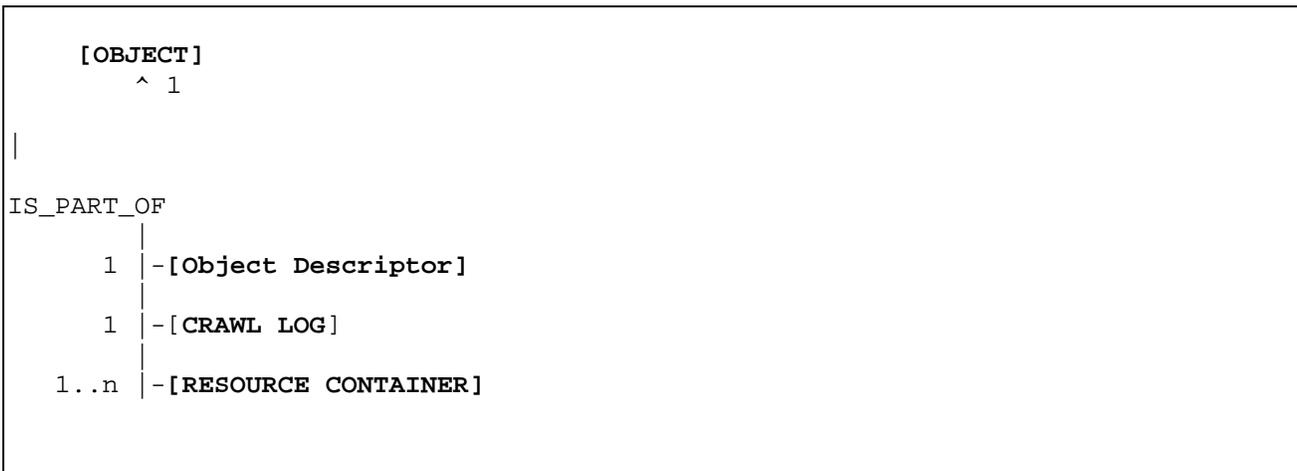
- **Delivery applications:** DRS Web Admin, Web Archiving Collection Service (WAX)
- **Rendering applications:** Web browser

Repository Notes

- **Assessment and plans:** No current assessments or plans

Content Model Details

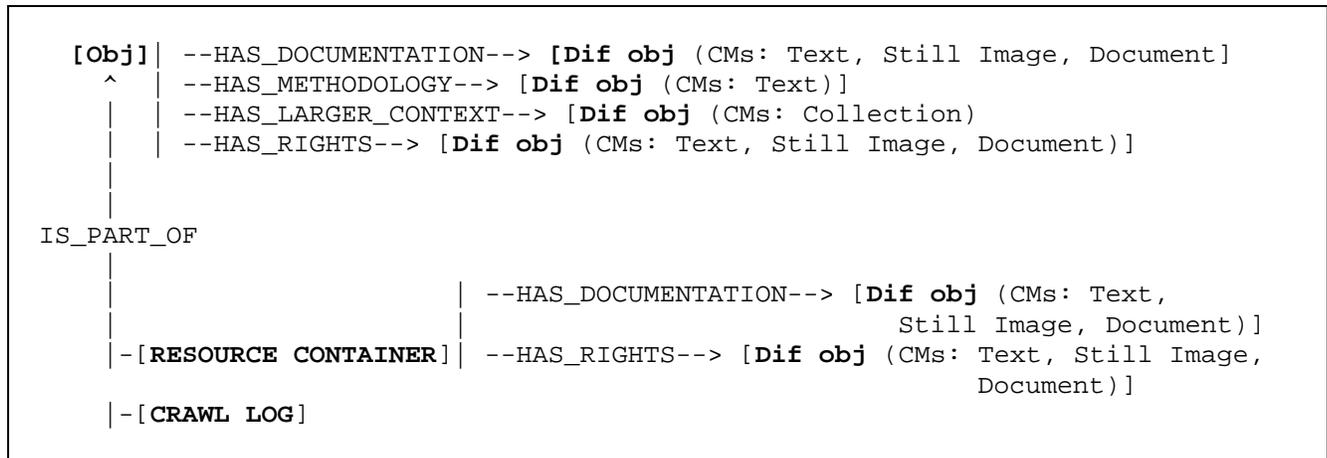
- **Structure:** a complex object manifest in multiple files: one or more resource container files containing crawled web resources and a crawl log documenting the process and outcome of the crawl.



Structure diagram for a web harvest object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationships between each child and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - No relationships
 - File-to-Different Object
 - Any resource container may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this

- object and its associated collection object
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a web harvest object, or the web harvest object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - The crawl log has a LOG role
 - Each resource container has a CONTAINER role
- **Directory structure**
 - The log file is in a directory called webharvest_log_1
 - The resource containers are in a directory called container_1

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- There won't be any URNs requested for this object's files during the automated deposit process.

Additional Metadata Requirements

Object component	Metadata
Resource container	<ul style="list-style-type: none"> ● Format name: GZIP <ul style="list-style-type: none"> ○ These will actually be one of: <ul style="list-style-type: none"> ▪ Compressed ARC files (with file extension arc.gz) ● MIME type: application/x-gzip ● Metadata type: - ● Usage class: HIGHUSE ● Access flag: B

	<ul style="list-style-type: none"> ● Role: CONTAINER
Crawl log	<ul style="list-style-type: none"> ● Format name: Plain Text ● MIME type: text/plain ● Metadata type: TEXT ● Charset: ASCII ● Usage class: LOWUSE ● Access flag: N ● Role: LOG

Descriptor Details

- METS element
 - TYPE attribute = “WEB HARVEST”
- Descriptive metadata in a MODS block
 - Seed metadata entered in WAXI
 - WAX Seed and harvest IDs
- Contains a seed list
 - During the migration this will be imported from the old METS
- Contains a harvestMD section (http://hul.harvard.edu/ois/xml/xsd/wax/harvestMD_1_0.xsd)
 - During the migration this will be imported from the old METS
- Contains the Heritrix crawl configuration (http://hul.harvard.edu/ois/xml/xsd/wax/heritrix_settings_1_x.xsd)

CMID-4.0 (PDS DOCUMENT) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** PDS DOCUMENT
- **Content Model ID:** CMID-4.0

Summary Description

- **Description:** This content model represents a page-based object delivered by Harvard Library's Page Delivery Service (PDS). The object is composed of multiple files, each representing a page image, page text and optionally the page text together with the layout (ALTO files).
- **Compatible object genres:** book, manuscript, pamphlet, photo album

Related Software and Systems

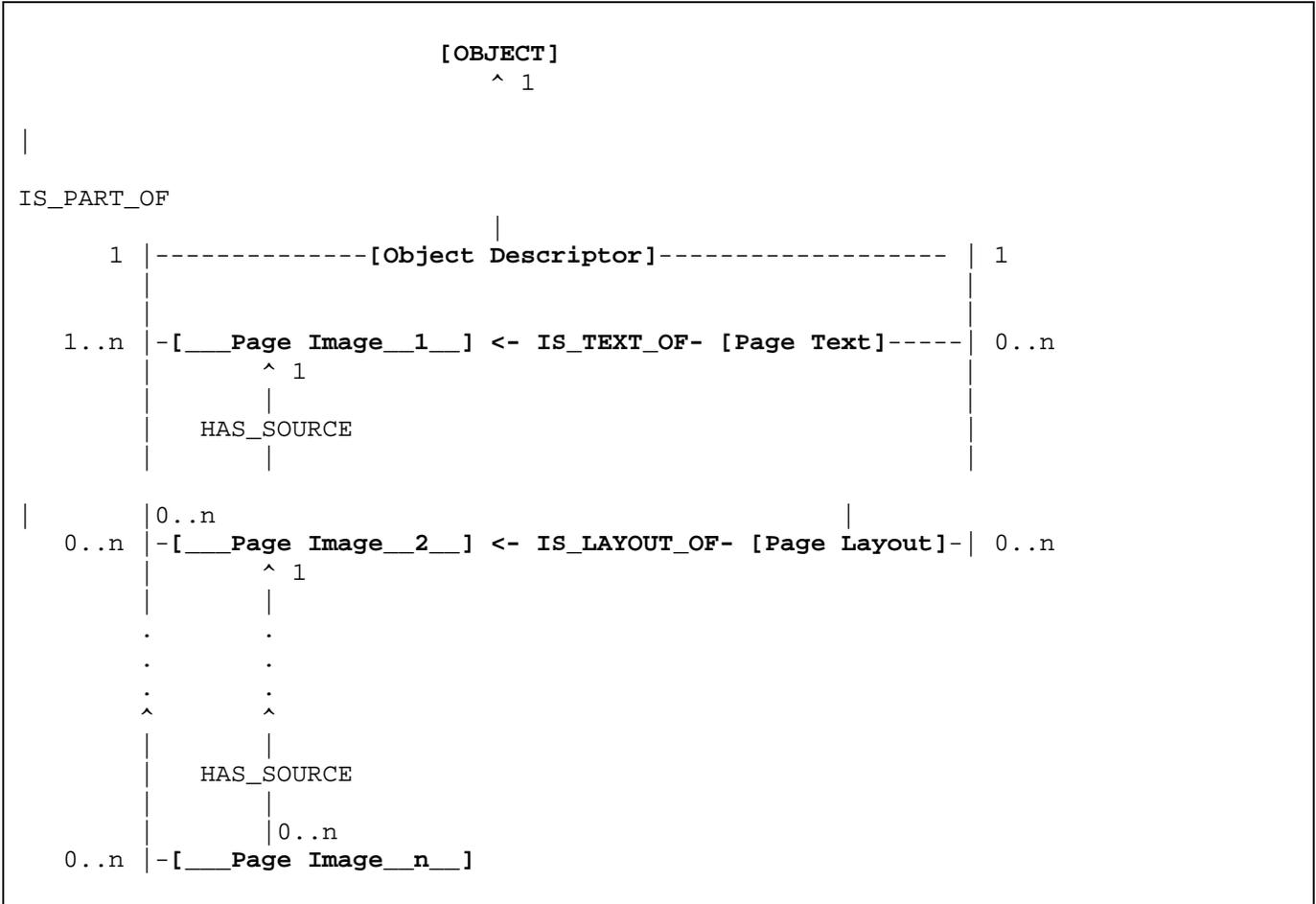
- **Delivery applications:** DRS Web Admin, Page Delivery Service (PDS), Harvard Viewer
 - Image Delivery Service (IDS) for the page images
- **Rendering applications:** Web browser

Repository Notes

- **Assessment and plans:** No current assessments or plans

Content Model Details

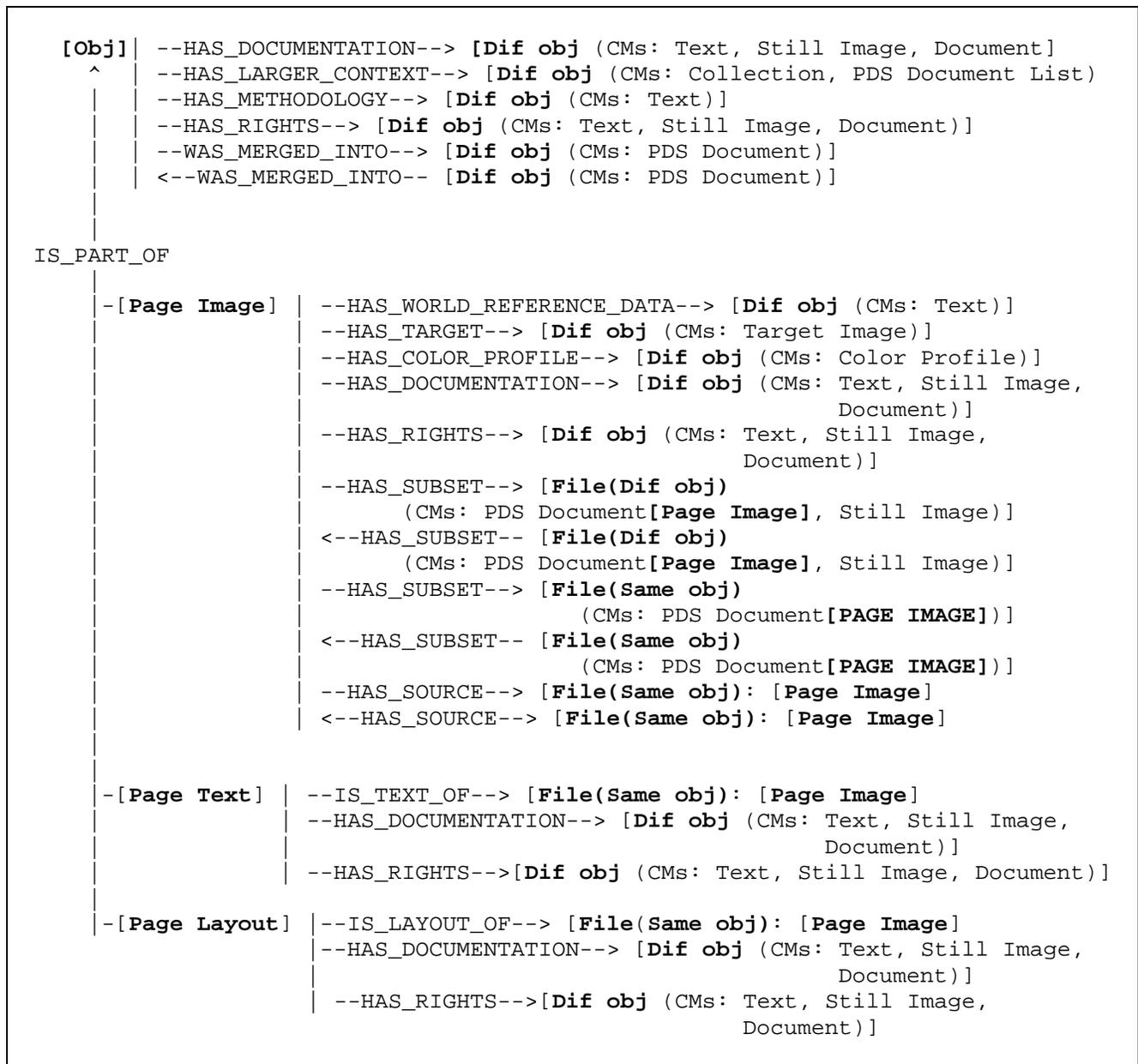
- **Structure:** a complex object manifest in multiple files: one or more page files, representing images, text or the page layout. The image files may have derivative images.



Structure diagram for a PDS document object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationships between each child and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_SOURCE relationships between derivative images
 - IS_TEXT_OF relationships between OCR text and images
 - These are implicitly modeled via the object descriptor's structMap by grouping text and images together
 - IS_LAYOUT_OF relationships between ALTO files and images
 - These are implicitly modeled via the object descriptor's structMap by grouping ALTO files and images together
 - Image files may be the source of HAS_SUBSET relationships between stitched images

- Image files may be the target of HAS_SUBSET relationships between stitched images
- File-to-Different Object
 - Image files may be the source of HAS_WORLD_REFERENCE_DATA relationships to objects conforming to the Text content model.
 - Image files may be the source of HAS_TARGET relationships to objects conforming to the Target Image content model.
 - Image files may be the source of HAS_COLOR_PROFILE relationships to objects conforming to the Color Profile content model.
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
- Object-to-Different Object
 - May be the source or target object of a WAS_MERGED_INTO relationship (valid only between objects of the PDS Document content model)
 - *May be the source object of an IS_PRESERVATION_REPLACEMENT relationship to a MOA2 document object (will remove this after the MOA2 deletions)*
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object or PDS Document List object
- File-to-File within Different Object
 - Image files may be the source of HAS_SUBSET relationships between stitched images
 - Image files may be the target of HAS_SUBSET relationships between stitched images



Relationship diagram showing what is possible in terms of relationships between files in a PDS document object, or the PDS document object itself, and other files or objects. (Key: Dif obj = Different object; File(Same obj) = File within the same object; CMs = Content Models)

- **Roles**

- Object
 - No roles
- Files
 - All page images have the role PAGE_IMAGE
 - All page text files have the role PAGE_TEXT
 - All page layout files have the role PAGE_COORDINATES

- It's not expected to be the case, but the files may also have one or more of the following roles:
 - ARCHIVAL_MASTER
 - DELIVERABLE
 - PRODUCTION_MASTER
- **Directory structure**
 - Directory structure will vary depending on the version of Batch Builder used.
 - Batch Builder 1
 - Images are contained in directories that begin with the prefix archival_master, production_master, or deliverable
 - Page text files are contained in directories that begin with the prefix ocr_corrected, ocr_uncorrected or keyed_text
 - Batch Builder 2
 - Images are contained in directories that begin with the prefix image
 - Page text files are contained in directories that begin with the prefix text
 - Page layout files are contained in directories that begin with the prefix layout

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: request URNs only for the content files having the role DELIVERABLE

Additional Metadata Requirements

Object component	Metadata
Object descriptor	<ul style="list-style-type: none"> ● Optional descriptive metadata in a MODS block ● PDS-specific metadata <ul style="list-style-type: none"> ○ http://hul.harvard.edu/ois/xml/xsd/pds/pds.xsd ● Page divisions in the structMap
Page Image	<ul style="list-style-type: none"> ● Format: per page image, one of the following: <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ▪ Format name: JPEG File Interchange Format ▪ MIME type: image/jpeg ○ 2 <ul style="list-style-type: none"> ▪ Format name: Graphics Interchange Format ▪ MIME type: image/gif ○ 3 <ul style="list-style-type: none"> ▪ Format name: Tagged Image File Format ▪ MIME type: image/tiff ○ 4 <ul style="list-style-type: none"> ▪ Format name: JPEG 2000 JP2 ▪ MIME type: image/jp2 ● Metadata type: IMAGE ● Role: PAGE_IMAGE <ul style="list-style-type: none"> ○ Page images may have one or more of the following roles: ARCHIVAL_MASTER, PRODUCTION_MASTER, DELIVERABLE

	<ul style="list-style-type: none"> ● Usage class: LOWUSE if the content file does not have the role DELIVERABLE; HIGHUSE if the content file has the role DELIVERABLE (suggested default)
Page Text	<ul style="list-style-type: none"> ● Format name: Plain Text ● MIME type: text/plain ● Metadata type: TEXT ● Usage class: HIGHUSE (suggested default) ● Role: PAGE_TEXT
(Page Text and) Layout	<ul style="list-style-type: none"> ● Format name: Extensible Markup Language <ul style="list-style-type: none"> ○ These are XML files using the ALTO schema ● MIME type: text/xml ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Role: PAGE_COORDINATES

Descriptor Details

- mets element
 - TYPE attribute = “PDS DOCUMENT”
 - May have a LABEL attribute
 - If it does this is used in PDS as the title
- structMap
 - structmap@TYPE = MIXED (was LOGICAL in the DRS 1 descriptors).
 - Within the structMap the div@TYPE is one of CITATION or PAGE
 - The sequence order (div@ORDER) must start with 1 and be in sequential order.

CMID-4.1 (DOCUMENT) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** DOCUMENT
- **Content Model ID:** CMID-4.1

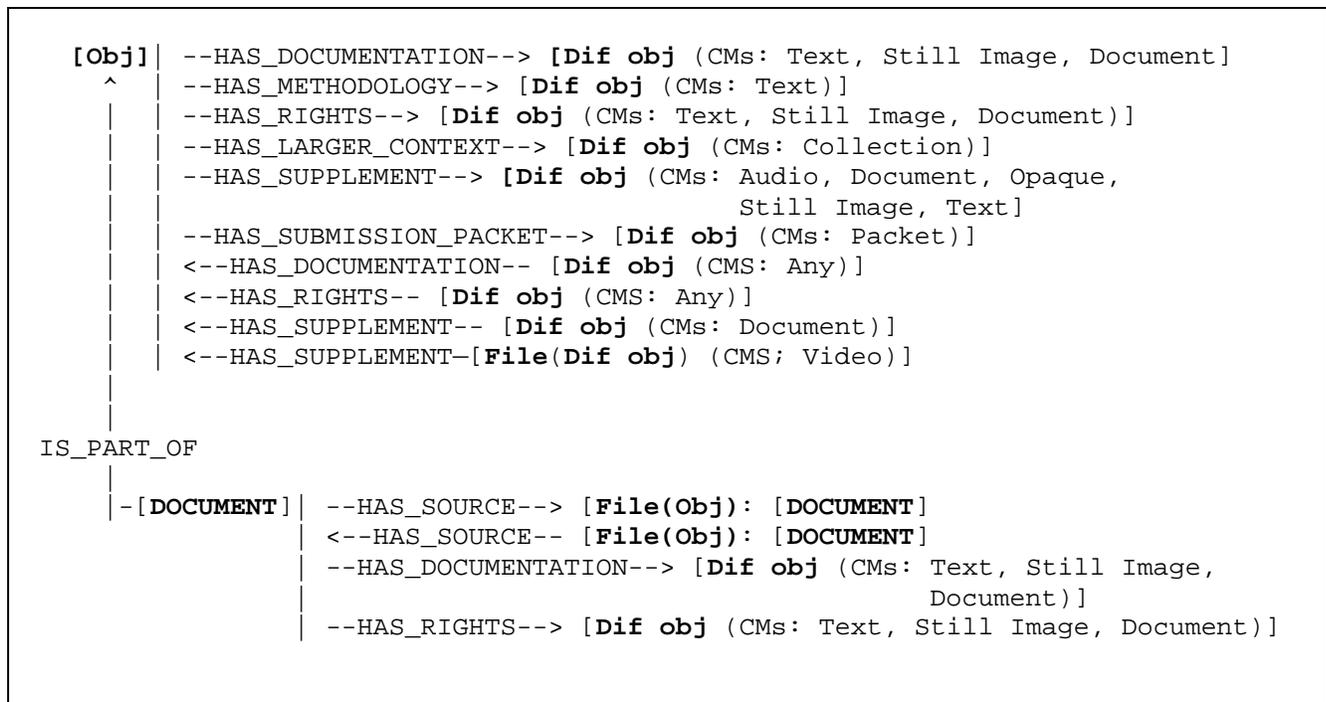
Summary Description

- **Description:** This content model represents a page-based object that can be delivered by HL's File Delivery Service (FDS). Each of the object's content files represents the entire page-based object. The object is composed primarily of text, but may include embedded multimedia (images, sounds, video etc) and specify fonts, colors, text size and backgrounds.
- **Compatible object genres:** book, manuscript, pamphlet, photo album, article, calendar, Email attachments

Related Software and Systems

- **Delivery applications:** DRS Web Admin, File Delivery Service (FDS)
- **Rendering applications:** Adobe Acrobat Reader or other PDF reader (for PDF), Microsoft Word

- File-to-its Object
 - IS_PART_OF relationships between each child and the object
 - These are implicitly modeled by describing the files in the object descriptor.
- File-to-File within the same Object
 - HAS_SOURCE relationships between derivative files
- File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODOLOGY, HAS_DOCUMENTATION, or HAS_SUPPLEMENT relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object (as in the case of email attachment objects)
 - May be the target object of HAS_RIGHTS, HAS_DOCUMENTATION, or HAS_SUPPLEMENT relationships from another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_SUBMISSION_PACKET relationship between this attachment object and its associated packet object.
- Object-to-File in Different object
 - The object may be the target of HAS_SUPPLEMENT relationships from video files in the video content model.
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a document object, or the document object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object (may have any of these roles)
 - DOCUMENTATION
 - DONOR_AGREEMENT
 - EMAIL_ATTACHMENT
 - EMBARGO_DOCUMENTATION
 - FINDING_AID
 - HARVARD_POLICY
 - LICENSE
 - RISK_ASSESSMENT
 - STATUTE
 - THESIS
 - THESIS_CERTIFICATE
 - THESIS_SUPPLEMENT
 - Files (may have any of the following roles)
 - ARCHIVAL_MASTER
 - DELIVERABLE
 - DOCUMENTATION
 - LICENSE
 - LOG
 - ORIGINAL_ORDER
 - PRODUCTION_MASTER

- **Directory structure**
 - The documents are contained within directories with the prefix document. Any subdirectories are also named with the prefix document.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: request a URN for each of this object's content files.

Additional Metadata Requirements

Object component	Metadata
Document	<ul style="list-style-type: none"> ● Format name: one of: <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ▪ Portable Document Format ▪ MIME media-type: application/pdf ○ 2 <ul style="list-style-type: none"> ▪ PDF/A ▪ MIME media-type: application/pdf ○ 3 <ul style="list-style-type: none"> ▪ PDF/X ▪ MIME media-type: application/pdf ○ 4 <ul style="list-style-type: none"> ▪ Microsoft Word Binary File Format ▪ MIME media-type: application/msword ○ 5 <ul style="list-style-type: none"> ▪ Office Open XML Document (.docx) ▪ MIME media-type: application/vnd.openxmlformats-officedocument.wordprocessingml.document ○ 6 <ul style="list-style-type: none"> ▪ Open Document Text ▪ MIME media-type: application/vnd.oasis.opendocument.text ○ 7 <ul style="list-style-type: none"> ▪ Rich Text Format (RTF) ▪ MIME media-type: application/rtf ○ 8 <ul style="list-style-type: none"> ▪ WordPerfect Document ▪ MIME media-type: application/vnd.wordperfect ○ 9 <ul style="list-style-type: none"> ▪ EPUB ▪ MIME media-type: application/epub+zip ● Metadata type: DOCUMENT ● Usage class: HIGHUSE (suggested default) ● Access flag: P (suggested default) ● Role: DELIVERABLE (suggested default)

Descriptor Details

- mets element

- TYPE attribute = “DOCUMENT”

CMID-4.3 (GOOGLE DOCUMENT CONTAINER 1) Status – IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** UNSUPPORTED

Identity

- **Aliases:** GOOGLE DOCUMENT CONTAINER 1 (preferred), GOOGLE 0.1 CONTAINER
- **Content Model ID:** CMID-4.3

Summary Description

- **Description:** This content model represents a page-based object created as part of the Google scanning project that could be delivered by HL's Page Delivery Service (PDS) if the object were brightened. The object's files are archived into a single compressed archive file. These objects were prepared and deposited by automated scripts written and maintained by LTS. This content model describes the first of three forms of the Google objects in which there was not an accompanying METS file.
- **Compatible object genres:** book

Related Software and Systems

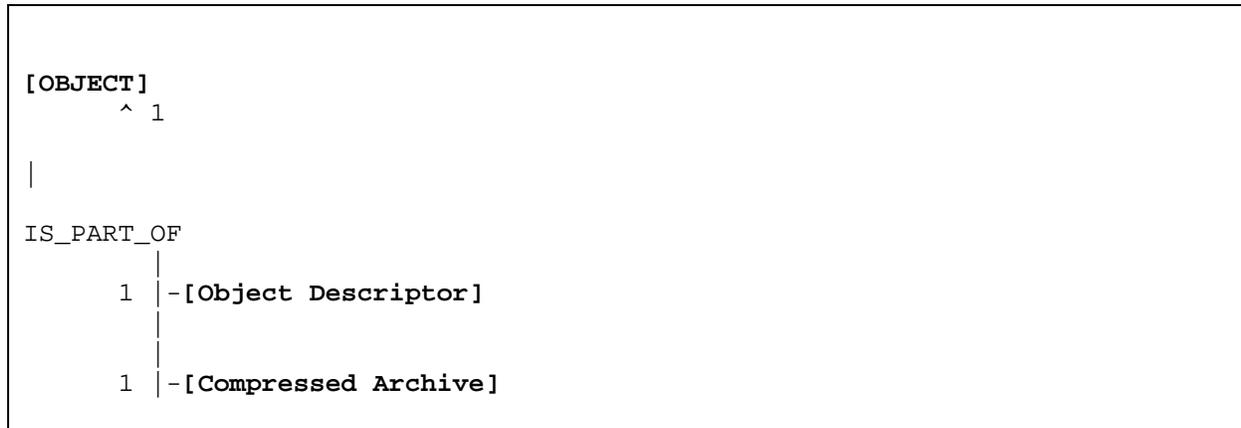
- **Delivery applications:** DRS Web Admin
- **Rendering applications:** none

Repository Notes

- **Assessment and plans:** This content model has been deprecated and has been superseded by CMID-4.4. HL attempted to re-deposit all of these books in the form of CMID-4.4 from the end of 2008 through early 2009, but 1,070 of these were not available for re-download from Google.

Content Model Details

- **Structure:** a single content file representing a scanned book



Structure diagram for a google document container 1 object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between the book container and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - No relationships
 - File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object
 - File-to-File within Different Object
 - No relationships

- Scanned Image Data
 - For each page image contains the calibration value (?), detected page number and page tags
 - Scanned Image Quality
 - Overall statistics on missing pages, etc.
 - One text file called pagedata.txt
 - For each page image contains the calibration value (?), detected page number and page tags
 - One text file called notes.txt
 - Contains quality metrics (estimated percentage of missing pages, etc.)
 - One text file called checksum.md5
 - Contains an MD5 for each file

Descriptor Details

- mets element
 - TYPE attribute = “GOOGLE DOCUMENT CONTAINER 1”

CMID-4.4 (GOOGLE DOCUMENT CONTAINER 2) *Status – IN-PRODUCTION*

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** UNSUPPORTED

Identity

- **Aliases:** GOOGLE DOCUMENT CONTAINER 2 (preferred), GOOGLE 1.0 CONTAINER
- **Content Model ID:** CMID-4.4

Summary Description

- **Description:** This content model represents a page-based object created as part of the Google scanning project that could be delivered by HL's Page Delivery Service (PDS) if the object were brightened. The object's files are archived into a single compressed file. These objects are prepared and deposited by automated scripts written and maintained by HL. This content model describes the second of three forms of the Google objects in which an accompanying METS file was introduced.
- **Compatible object genres:** book

Related Software and Systems

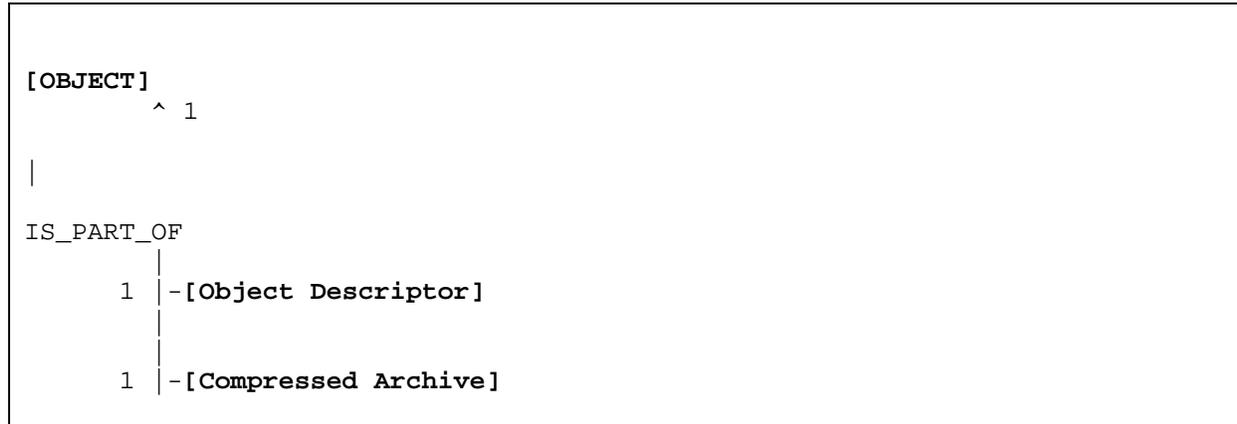
- **Delivery applications:** DRS Web Admin
- **Rendering applications:** none

Repository Notes

- **Assessment and plans:** During the migration we will see if we can migrate this and the other 2 variations (CMID-4.3 and CMID-4.6) to the same format.

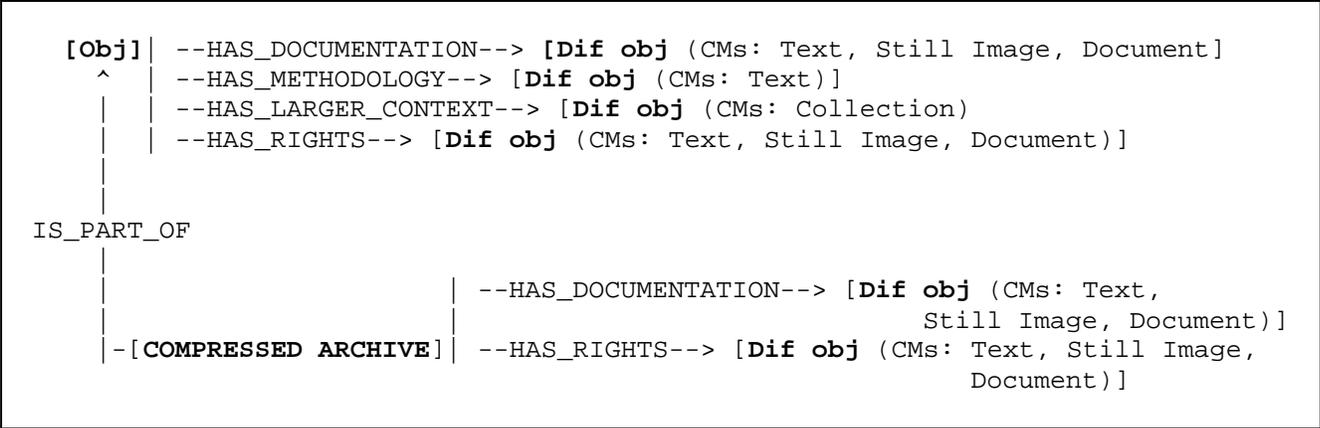
Content Model Details

- **Structure:** a single content file representing a scanned book



Structure diagram for a google document container 2 object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between the book container and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - No relationships
 - File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object
 - File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a google document container 2 object, or the google document container 2 object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - The compressed archive will have the role CONTAINER

- **Directory structure**
 - The container file is not under any subdirectories. Within the container file there are no subdirectories.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: do not request a URN for the content file

Additional Metadata Requirements

Object component	Metadata
Book container	<ul style="list-style-type: none"> ● Format name: ZIP ● MIME type: application/zip ● Metadata type: - ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: CONTAINER (suggested default) <ul style="list-style-type: none"> ○ Within this container the following bitstreams (filestreams) are valid: <ul style="list-style-type: none"> ▪ one or more JP2 page images ▪ one or more plain text files for OCR of page images ▪ one METS XML file <ul style="list-style-type: none"> ● The name of the file is HARVARD_*.xml where * is a sequence of numbers and/or letters (a barcode?) <ul style="list-style-type: none"> ○ In a sample object the name is

HARVARD_FL4IKQ.xml

- Contains production notes (bad, missing pages, etc.), tags used in the structmap, Premis events for the scan, processing, analysis, conversion of deliverables

Descriptor Details

- mets element
 - TYPE attribute = “GOOGLE DOCUMENT CONTAINER 2”

CMID-4.6 (GOOGLE DOCUMENT CONTAINER 3) Status – IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** UNSUPPORTED

Identity

- **Aliases:** GOOGLE DOCUMENT CONTAINER 3 (preferred), GOOGLE 1.1 CONTAINER
- **Content Model ID:** CMID-4.6

Summary Description

- **Description:** This content model represents a page-based object created as part of the Google scanning project that could be delivered by HL's Page Delivery Service (PDS) if the object were brightened. The object's files are archived into a single compressed file. These objects are prepared and deposited by automated scripts written and maintained by HL. This content model describes the third of three forms of the Google objects in which changes were made to the accompanying METS file.
 - This content model is very similar to CMID-4.4 (Google Document Container 2, Google 1.0 Container). The differences are that the METS file in this content model includes a library identifier, and the identifier embedded within the JPEG 2000 images uses a different character to separate values.
- **Compatible object genres:** book

Related Software and Systems

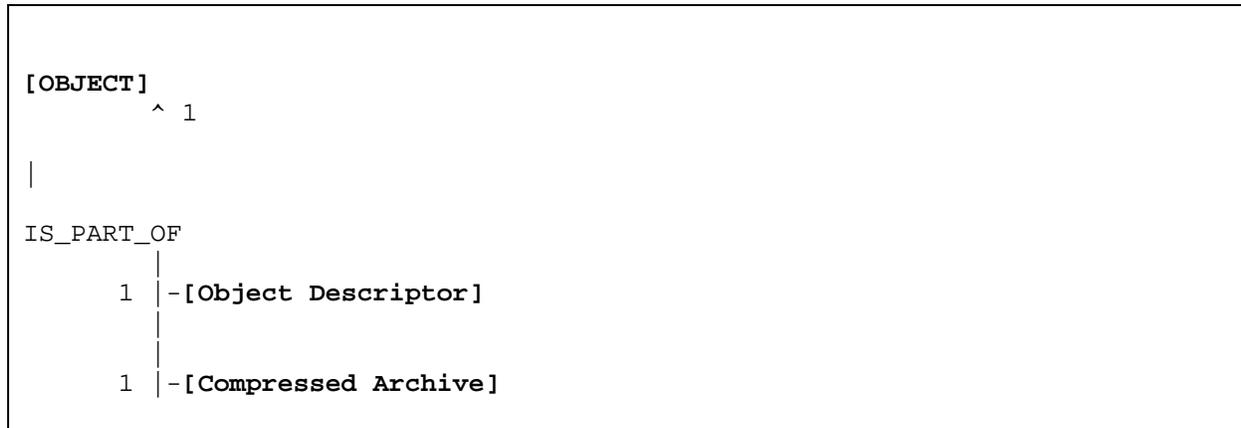
- **Delivery applications:** DRS Web Admin
- **Rendering applications:** none

Repository Notes

- **Assessment and plans:** During the migration we will see if we can migrate this and the other 2 variations (CMID-4.3 and CMID-4.4) to the same format.

Content Model Details

- **Structure:** a single content file representing a scanned book



Structure diagram for a google document container 3 object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between the book container and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - No relationships
 - File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object
 - File-to-File within Different Object
 - No relationships

- Contains a MARC record, production notes (bad, missing pages, etc.), tags used in the structmap, barcode, “significant properties” (file count, number of pages), Premis events for the scan, processing, analysis, conversion of deliverables

Descriptor Details

- mets element
 - TYPE attribute = “GOOGLE DOCUMENT CONTAINER 3”

CMID-5.0 (STILL IMAGE) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** STILL IMAGE
- **Content Model ID:** CMID-5.0

Summary Description

- **Description:** This content model represents visual works in digital form. Each object represents a single image (photograph, map, etc.).
- **Compatible object genres:** photograph, map, Email attachment

Related Software and Systems

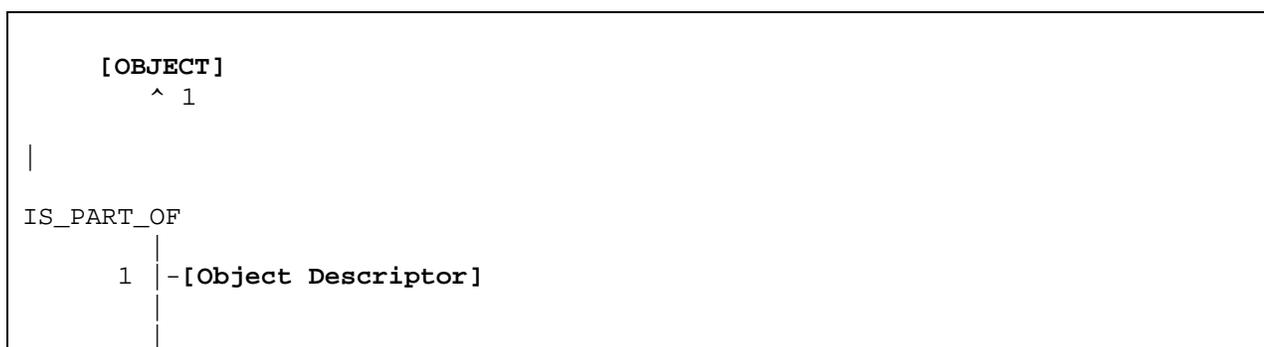
- **Delivery applications:** DRS Web Admin, IDS (TIFF, JPEG, JP2, GIF formats only)
 - The JP2 images are converted to JPEG format before delivery by IDS
- **Rendering applications:** Web browser (JPEG, GIF), Image viewer or Web browser plug-in (TIFF)

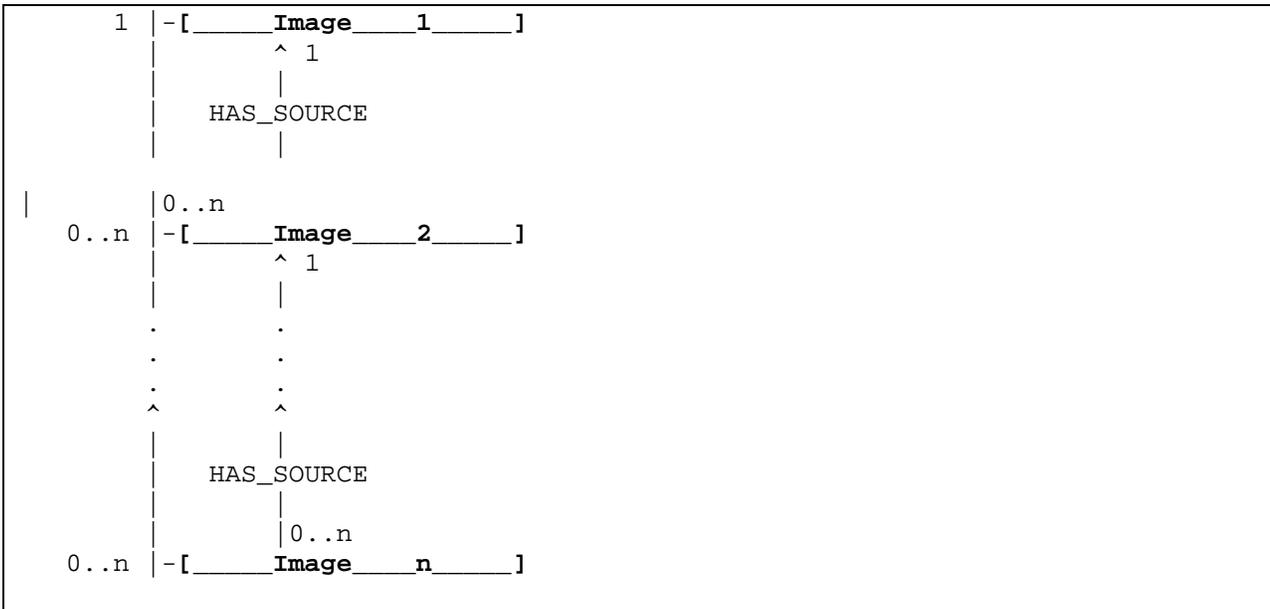
Repository Notes

- **Assessment and plans:** No current assessments or plans

Content Model Details

- **Structure:** A single parent image, possibly related to derivative images.



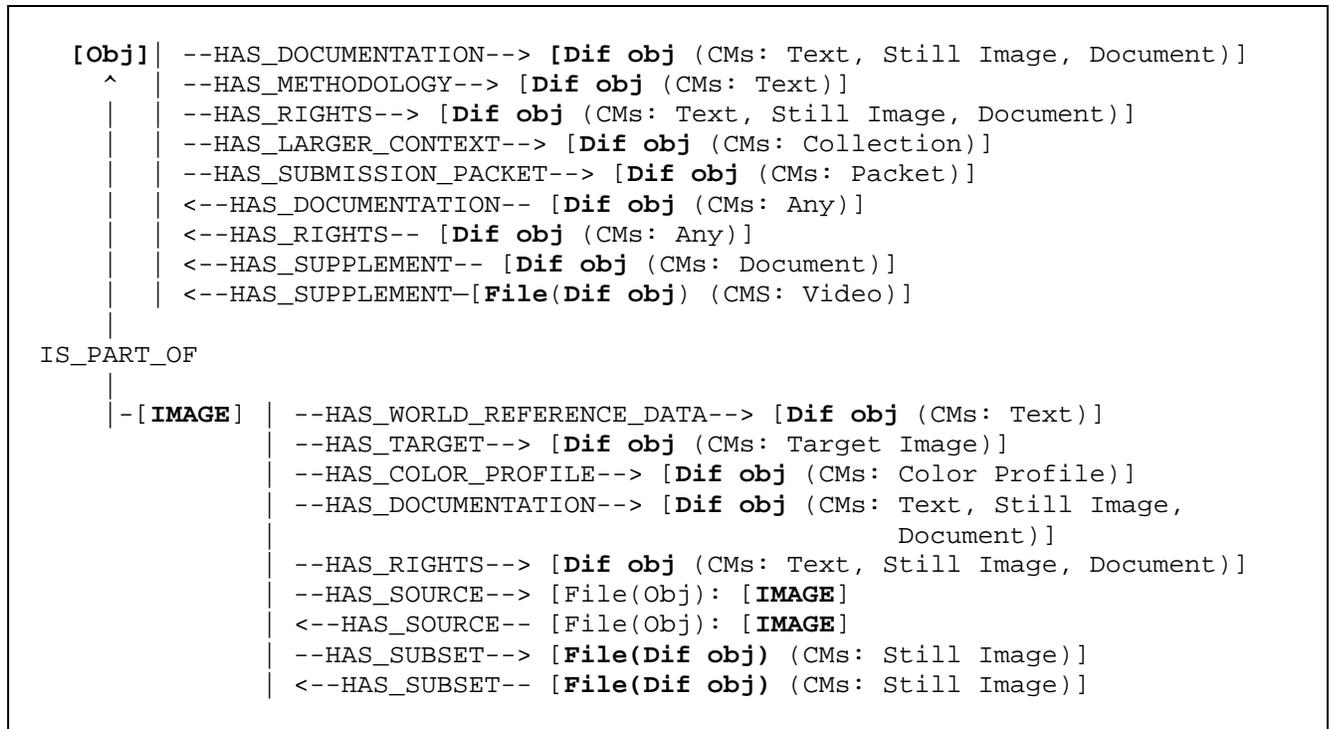


Structure diagram for a still image object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationships between each file and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_SOURCE relationships between derivative images
 - File-to-Different Object
 - Image files may be the source of HAS_WORLD_REFERENCE_DATA relationships to objects conforming to the Text content model.
 - Image files may be the source of HAS_TARGET relationships to objects conforming to the Target Image content model.
 - Image files may be the source of HAS_COLOR_PROFILE relationships to objects conforming to the Color Profile content model.
 - Image files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object (as in the case of email attachment objects)
 - May be the target object of HAS_RIGHTS, HAS_DOCUMENTATION, or HAS_SUPPLEMENT relationships from another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor

descriptor

- May be the source object of a HAS_SUBMISSION_PACKET relationship between this attachment object and its associated packet object.
- Object-to-File in Different object
 - The object may be the target of HAS_SUPPLEMENT relationships from video files in the video content model.
- File-to-File within Different Object
 - Image files may be the source of HAS_SUBSET relationships between stitched images
 - Image files may be the target of HAS_SUBSET relationships between stitched images



Relationship diagram showing what is possible in terms of relationships between files in a still image object, or the still image object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models; File(obj) = File within the same object)

- **Roles**

- Object (may have one or more of the following roles)
 - DOCUMENTATION
 - DONOR_AGREEMENT
 - EMAIL_ATTACHMENT
 - FINDING_AID
 - HARVARD_POLICY
 - LICENSE
 - RISK_ASSESSMENT
 - STATUTE
 - THESIS_SUPPLEMENT
- Files (may have one or more of the following)
 - ARCHIVAL_MASTER

- DELIVERABLE
- DOCUMENTATION
- LICENSE
- PRODUCTION_MASTER
- THUMBNAIL

Images within this content model are differentiated from each other primarily through their roles. Any images derived from the archival master image that have been processed (color correction, cropping, etc.) should be designated with the role PRODUCTION_MASTER. Any images that are intended to be use copies should be designated with the DELIVERABLE role. Note that images may have more than one role, such as in the case where an image is both the archival master and the use copy.

- **Directory structure**

- Directory structure will vary depending on the version of Batch Builder used.
 - Batch Builder 1
 - Images are contained in directories that begin with the prefix archival_master, production_master, or deliverable
 - Batch Builder 2
 - Images are contained in directories that begin with the prefix image. Any subdirectories are also named with the prefix image.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: request URNs only for the content files having the role DELIVERABLE

Additional Metadata Requirements

Object component	Metadata
Image	<ul style="list-style-type: none"> ● Format: per image, one of the following: <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ▪ Format name: JPEG File Interchange Format ▪ MIME type: image/jpeg ○ 2 <ul style="list-style-type: none"> ▪ Format name: Graphics Interchange Format ▪ MIME type: image/gif ○ 3 <ul style="list-style-type: none"> ▪ Format name: Tagged Image File Format ▪ MIME type: image/tiff ○ 4 <ul style="list-style-type: none"> ▪ Format name: JPEG 2000 JP2 ▪ MIME type: image/jp2 ○ 5 (This one is not accepted for new deposits but DRS1 objects may have this format) <ul style="list-style-type: none"> ▪ Format name: Kodak PhotoCD Image ▪ MIME type: image/x-photo-cd ● Metadata type: IMAGE

- Role: one or more of:
 - ARCHIVAL_MASTER
 - Usage class: LOWUSE (suggested default)
 - Access flag: N (suggested default)
 - PRODUCTION_MASTER
 - Usage class: LOWUSE (suggested default)
 - Access flag: N (suggested default)
 - DELIVERABLE
 - Usage class: HIGHUSE (suggested default)
 - Access flag: P (suggested default)
 - THUMBNAIL
 - Usage class: HIGHUSE (suggested default)
 - Access flag: P (suggested default)
 - DOCUMENTATION
 - Usage class: LOWUSE (suggested default)
 - Access flag: P (suggested default)
 - LICENSE
 - Usage class: LOWUSE (suggested default)
 - Access flag: P (suggested default)
- One image in this object should have an isFirstGenerationInDRS value of true
- One image in this object should have an isPreferredDeliverableSource value of true

CMID-5.2 (TARGET IMAGE) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** TARGET IMAGE
- **Content Model ID:** CMID-5.2

Summary Description

- **Description:** This content model represents a target image for one or more other images. Targets document the setup and calibration of digital reformatting processes.
- **Compatible object genres:** target

Related Software and Systems

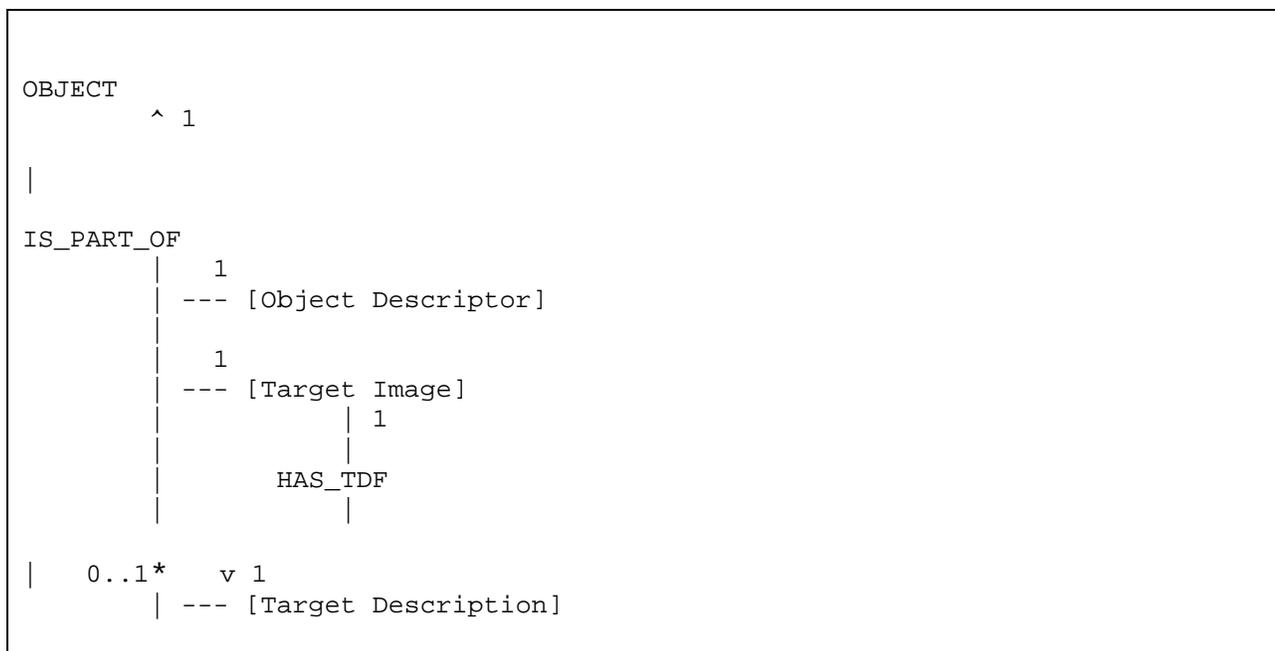
- **Delivery applications:** DRS Web Admin, IDS (TIFF, JPEG, JP2, GIF formats only)
 - The JP2 images are converted to JPEG format before delivery by IDS
- **Rendering applications:** Web browser (JPEG, GIF), Image viewer or Web browser plug-in (TIFF)

Repository Notes

- **Assessment and plans:** No current assessments or plans

Content Model Details

- **Structure:** a single image file representing a target with an optional textual description of the target image

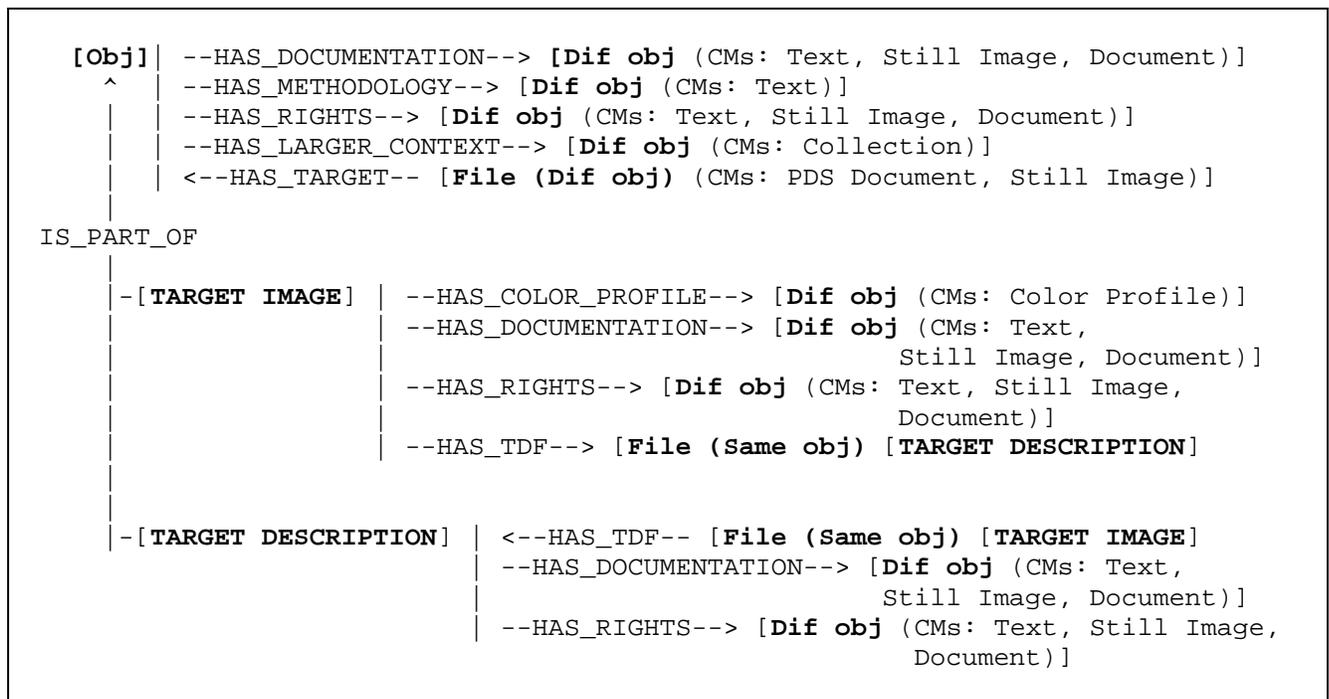


Structure diagram for a target image object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

*** There is one instance in the DRS where a target image file (DRS file ID 246628) is related to 3 target description files (DRS file IDs 246629, 246625, 246626). This is because the image contains captures of 3 different targets. It is best practice to deposit and describe these separately, so at most 1 target description per target image file is supported for DRS ingest, although this case exists in the DRS.**

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between the each file and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_TDF relationship from the target image to the target description
 - File-to-Different Object
 - Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - Target image files may be the source of HAS_COLOR_PROFILE relationships to objects in the Color Profile content model.
 - This object may be the target of HAS_TARGET relationships from image files in the Still

- Image or PDS Document content model
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a target image object, or the target image object itself, and other files or objects. (Key: Dif obj = Different object; File (Same obj) = File within the same object; File (Dif obj) = File within a different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - TARGET_IMAGE role for the target image
 - TARGET_DESCRIPTION role for the target description file
- **Directory structure**
 - Image files are located within a directory having the prefix image.
 - /image
 - Target description files are located in a directory having the prefix TDF.

- /tdf

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: do not request a URN for the content file

Additional Metadata Requirements

Object component	Metadata
Target image	<ul style="list-style-type: none"> ● Format <ul style="list-style-type: none"> ○ The following formats are supported: <ul style="list-style-type: none"> ▪ TIFF (MIME type: image/tiff) ▪ JPEG (MIME type: image/jpeg) ▪ JP2 (MIME type: image/jp2) ▪ JPX (MIME type: image/jpx) ▪ GIF (MIME type: image/gif) ▪ PhotoCD (MIME type: image/photo-cd) ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: TARGET_IMAGE
Target description	<ul style="list-style-type: none"> ● Format: Plain Text (MIME type: text/plain) ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default)

CMID-6.0 (TEXT) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** TEXT
- **Content Model ID:** CMID-6.0

Summary Description

- **Description:** This content model represents a single relatively simple text file (potentially with derivative files). They do not contain images, audio, video or support interactive behavior. They do not support custom fonts, colors, background images, etc. They can contain hypertext links or references to external files or objects.
- **Compatible object genres:** schema, OCR text, finding aids (EADs), license agreements, documentation, world files, Email attachments, audio playlists

Related Software and Systems

- **Delivery applications:** DRS Web Admin, File Delivery Service (FDS)
- **Rendering applications:** Web browser, general-purpose text viewer or editor (WordPad,

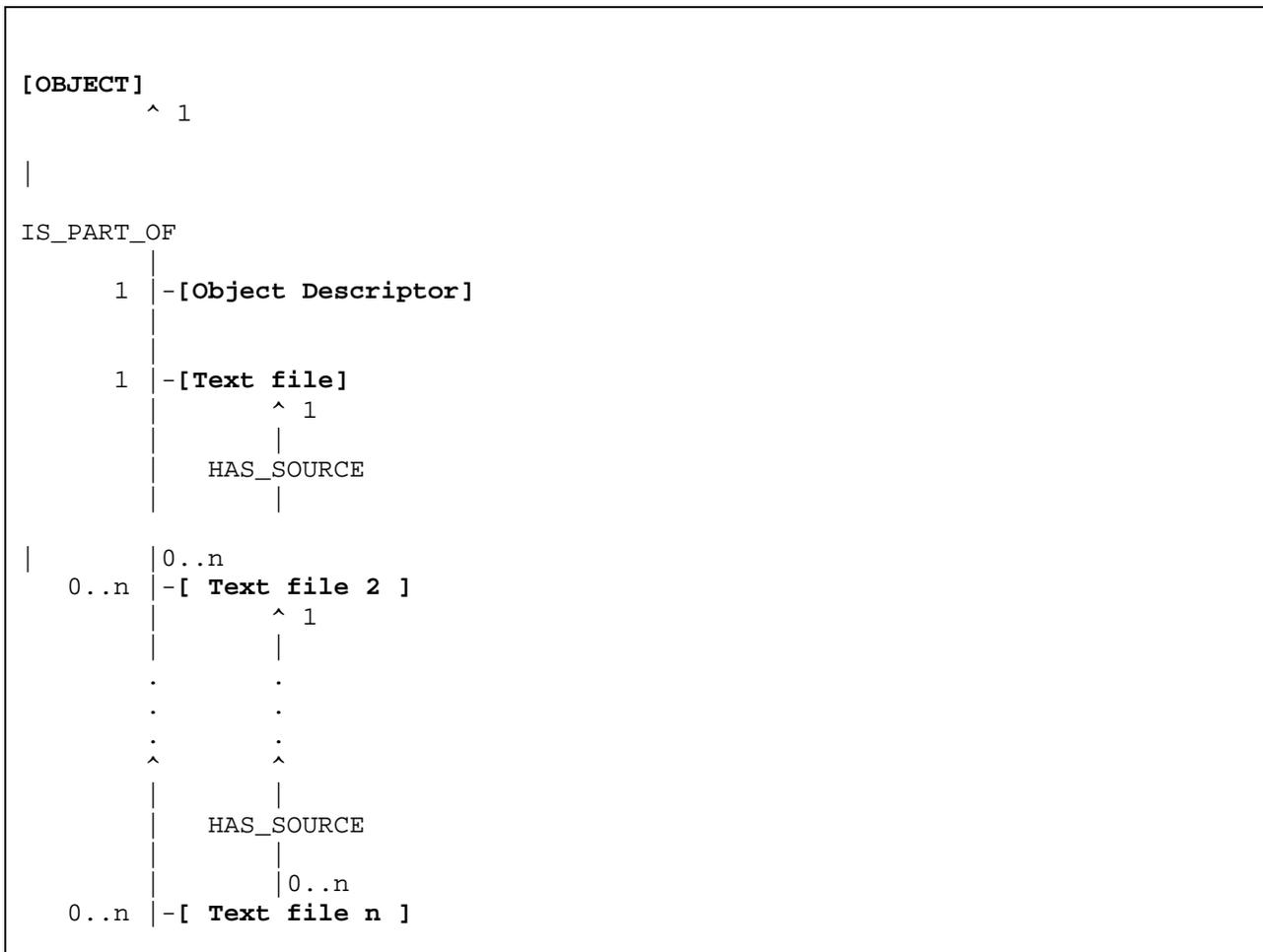
TextPad, vi, emacs, etc.), or format-specific (XML, SGML) editor

Repository Notes

- **Assessment and plans:** No current assessments or plans

Content Model Details

- **Structure:** a single text file with optional derivative files. In the future text components may include additional files associated with text files (schema, etc.).



Structure diagram for a text object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between each file and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_SOURCE relationships between derivative files
 - File-to-Different Object
 - The text file may be the source of HAS_RIGHTS or HAS_DOCUMENTATION

- **Roles**
 - Object (may have one or more of the following)
 - CLOSED_CAPTION_DATA
 - CODE
 - DOCUMENTATION
 - DONOR_AGREEMENT
 - EMAIL_ATTACHMENT
 - FINDING_AID
 - HARVARD_POLICY
 - LICENSE
 - PRESENTATION
 - RISK_ASSESSMENT
 - SCHEMA
 - STATUTE
 - THESIS_SUPPLEMENT
 - WORLD_REFERENCE_DATA
 - The text file may optionally have one or more of the following roles
 - DOCUMENTATION
 - LICENSE
 - LOG
 - ORIGINAL_ORDER
 - STYLE_SHEET

- **Directory structure**
 - Text files are located within a directory having the prefix text.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: request a URN for the text file

Additional Metadata Requirements

Object component	Metadata
Text file	<ul style="list-style-type: none"> ● Format <ul style="list-style-type: none"> ○ The following formats are supported: <ul style="list-style-type: none"> ▪ Any text format (MIME media-type: text/*) ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Access flag: P (suggested default)

CMID-8.0 (COLOR PROFILE) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** COLOR PROFILE
- **Content Model ID:** CMID-8.0

Summary Description

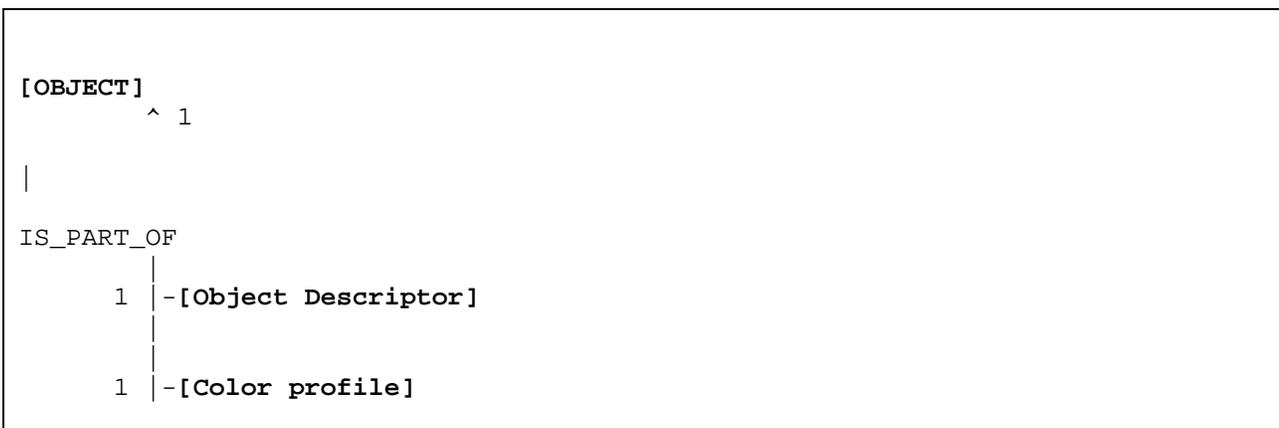
- **Description:** This content model represents color profiles which are used to translate color data created on one device into another device's native color space.
- **Compatible object genres:** image color profiles

Related Software and Systems

- **Delivery applications:** DRS Web Admin, FDS
- **Rendering applications:** N/A. Color profiles aren't rendered per se but can be used by imaging applications to render images. There are third party tools that can make the content of ICC color profiles readable, such as ICC Profile Inspector (<http://www.color.org/profileinspector.xalter>).

Repository Notes

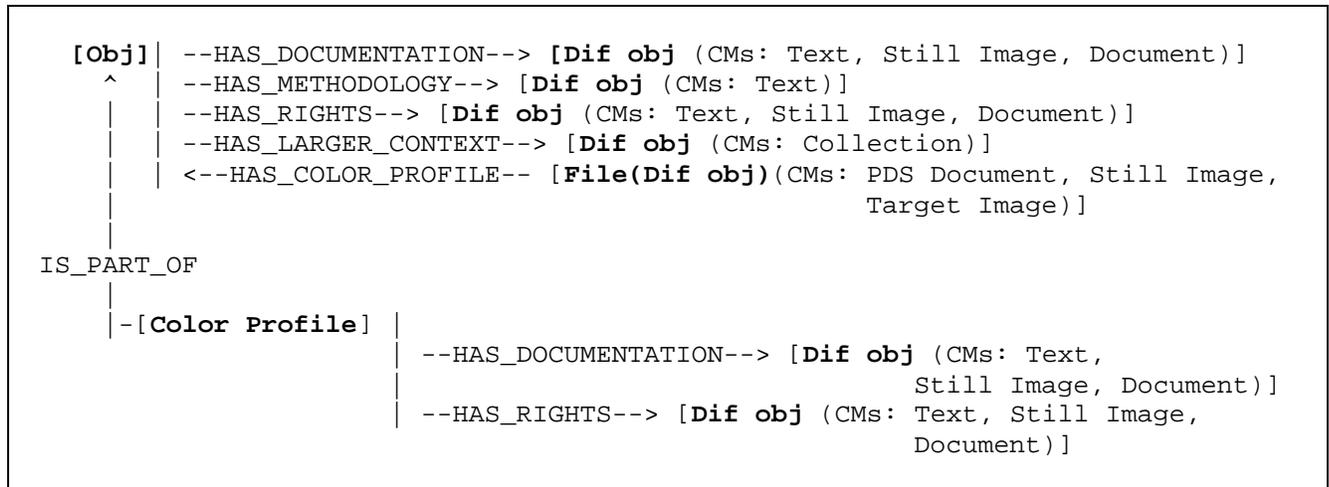
- **Assessment and plans:** No current assessments or plans
- **Content Model Details**
- **Structure:** a single color profile.



Structure diagram for a color profile object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - File-to-its Object
 - IS_PART_OF relationship between each file and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - No relationships
 - File-to-Different Object
 - The color profile file may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the target object of HAS_COLOR_PROFILE relationships from an image file in a PDS Document, Still Image or Target Image object.

- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a still image object, or the still image object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models; File(obj) = File within the same object)

- **Roles**
 - Object (may have one of the following)
 - COLOR_PROFILE
 - The text file may optionally have one or more of the following roles
- **Directory structure**
 - The color profile file is located within a directory having the prefix color_profile.
 - /color_profile

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: no URN for the color profile file

Additional Metadata Requirements

Object component	Metadata
Color profile	<ul style="list-style-type: none"> ● Format: ICC File Format for Color Profiles (MIME type: application/x-icc) ● Metadata type: N/A

- Usage class: LOWUSE (suggested default)
- Access flag: P (suggested default)

CMID-9.0 (EMAIL MESSAGE) Status - IN-PRODUCTION

Definition status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** EMAIL MESSAGE (preferred), EMAIL
- **Content model ID:** CMID-9.0

Summary description

- **Description:** This content model represents a single email message.
- **Compatible object genres:** email message

Related software and systems

- **Delivery applications:** DRS Web Admin
- **Rendering applications:** TBD

Assessments and preservation plans

- No current assessments or plans

Content Model Details

- **Structure:** This object contains a single email message file that does not have any in-line files or attachments. If the original email message had in-line files or attachments, they were stripped out of this message prior to DRS deposit.
 - Note that the in-line files and attachments exist in the DRS as separate related objects.
- Email that was processed in the Electronic Archiving System (EAS) will have associated pre-DRS process history in the descriptor.

```

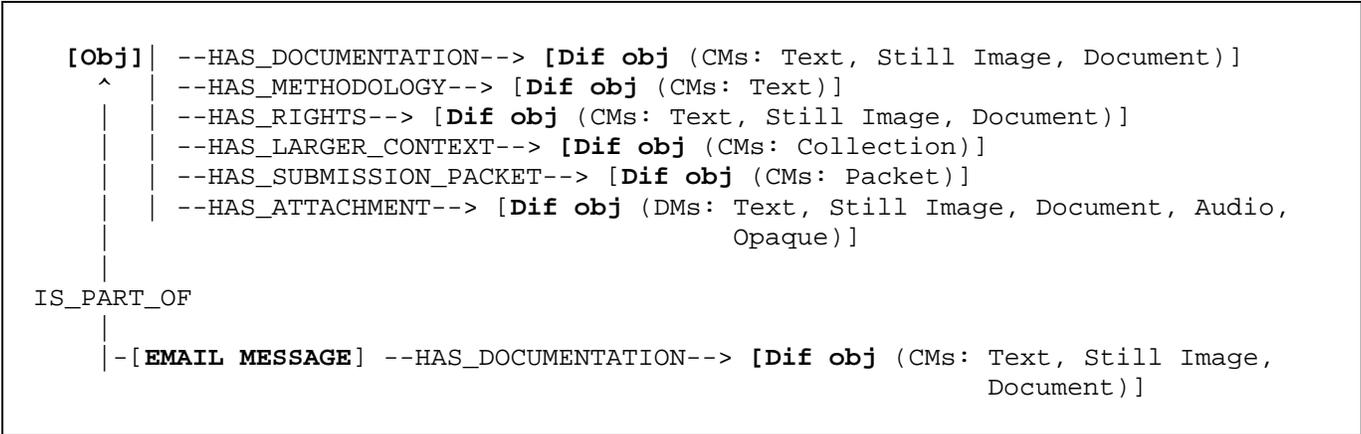
[OBJECT]
  ^ 1
  |
  | IS_PART_OF
  |
  1 |-[Object Descriptor]
  |
  1 |-[Email Message without in-line or attached files]

```

Structure diagram for an email message object showing what is possible in terms of number of files and relationships within the object, assuming all of the content files have not been deleted.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)

- File-to-its Object
 - IS_PART_OF relationship between the child files and the object
 - This is implicitly modeled by describing the files in the object descriptor.
- File-to-File within the same Object
 - No relationships
- File-to-Different Object
 - The email message file may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in the descriptor in PREMIS Rights metadata
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object.
 - May be the source object of a HAS_SUBMISSION_PACKET relationship between this object and its associated packet object.
 - May be the source object of a HAS_ATTACHMENT relationship between this object and an email attachment object in the Opaque, Audio, Document, Text or Still Image content models.
- File-to-File within Different Object
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in an email message object, or the email message object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- Roles
 - Object
 - No roles
 - Files
 - No roles

- **Directory structure**
 - There's a single top-level directory:
 - message
 - This directory contains the email message

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.
- Suggested default: no URNs for the files

Metadata Requirements

Object component	Metadata requirements
Email message (without in-line or attached files)	<ul style="list-style-type: none"> ● Format, one of: <ul style="list-style-type: none"> ○ RFC 822 mail text <ul style="list-style-type: none"> ▪ The format is EML with message files conforming to RFC-2822 but the file utility wrapped by FITS will report this format as RFC 822 mail text, which is technically still correct. ▪ If the message had attachments, they were stripped out to form separate related DRS objects before depositing the message into the DRS. ▪ The file extension is .eml ▪ MIME type: message/rfc822 ▪ Format technical metadata: none ○ Plain Text and Text variations (HTML, x-mail, pascal, etc.) (For email originating from Outlook) <ul style="list-style-type: none"> ▪ The format is EML with message files conforming to RFC-2822 but the file utility wrapped by FITS will report this format as Plain Text or a text variation if the email originated from Outlook. ▪ If the message had attachments, they were stripped out to form separate related DRS objects before depositing the message into the DRS. ▪ The file extension is .eml ▪ MIME type: text/* ▪ Format technical metadata: TextMD ● File metadata <ul style="list-style-type: none"> ○ Role: none ○ Usage class: LOWUSE (suggested default) ○ Access flag: N (suggested default)

Descriptor Details

- mets element
 - TYPE attribute = "EMAIL MESSAGE"
- In the administrative metadata (hulDrsAdmin schema), the removal of admin flags in EAS is recorded as versioned metadata
- In addition to the amdSec sections common to all content files:
 - May have one or more PREMIS events with metadata including the EAS event ID in the

- hulEventExtension schema
- structMap section
 - 1 div section

CMID-10.0 (COLLECTION) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** COLLECTION
- **Content Model ID:** CMID-10.0

Summary Description

- **Description:** This content model represents a collection of content.
- **Compatible object genres:** email account, email collection

Related Software and Systems

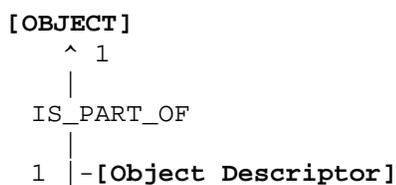
- **Delivery applications:** DRS Web Admin
- **Rendering applications:** TBD

Repository Notes

- **Assessment and plans:** No plans or assessments

Content Model Details

- **Structure:** This object does not have content files; it just has an object descriptor.



Structure diagram for a collection object showing what is possible in terms of number of files and relationships within the object.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - F-SO (File-to-Same Object)
 - N/A
 - F-F(SO) (File-to-File within Same Object)
 - N/A
 - F-DO (File-to-Different Object)
 - N/A

- O-O (Object-to-Object)
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor rather than as PREMIS relationships.
 - May be the target object of a HAS_LARGER_CONTEXT relationship between email and attachment objects and this object.
- F-F(DO) (File-to-File within Different Object)
 - No relationships

```
[Obj] | --HAS_DOCUMENTATION--> [Dif obj (CMs: Text, Still Image, Document)]
      | --HAS_METHODODOLOGY--> [Dif obj (CMs: Text)]
      | --HAS_RIGHTS--> [Dif obj (CMs: Text, Still Image, Document)]
      | <--HAS_LARGER_CONTEXT-- [Dif obj (CMs: Email, Text, Still Image, Document,
      |                               Audio, Opaque)]
```

Relationship diagram showing what is possible in terms of relationships between files in a collection object, or the collection object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - No roles
- **Directory structure**
 - Because there aren't any content files, there aren't any directories.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.

Metadata Requirements

- Because there aren't any content files, this section does not apply.

Descriptor Details

- mets element
 - TYPE attribute = "COLLECTION"
- No fileSec section
- A structMap section with a single empty div

CMID-11.0 (PACKET) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** PACKET
- **Content Model ID:** CMID-11.0

Summary Description

- **Description:** This content model represents a set of pre-processed content packaged, transferred and/or submitted together to a system other than the DRS. Currently these packets are limited to those submitted to Harvard’s Electronic Archiving System (EAS) but in the future could be expanded to apply to content submitted to other systems.
- **Compatible object genres:** an email account in its native email client format

Related Software and Systems

- **Delivery applications:** DRS Web Admin
- **Rendering applications:** N/A

Repository Notes

- **Assessment and plans:** No plans or assessments

Content Model Details

- **Structure:** This object does not have content files; it just has an object descriptor. If the packet was submitted to the Electronic Archiving System (EAS), the descriptor will have associated process history about the packet’s content, particularly content that was included in the packet that was deleted in EAS.



Structure diagram for a packet object showing what is possible in terms of number of files and relationships within the object.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - F-SO (File-to-Same Object)
 - N/A
 - F-F(SO) (File-to-File within Same Object)
 - N/A
 - F-DO (File-to-Different Object)
 - N/A
 - O-O (Object-to-Object)
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor rather than as PREMIS relationships.

- May be the target object of a HAS_SUBMISSION_PACKET relationship between email and attachment objects and this object.
- F-F(DO) (File-to-File within Different Object)
 - N/A

```
[Obj] | --HAS_DOCUMENTATION--> [Dif obj (CMs: Text, Still Image, Document)]
      | --HAS_METHODODOLOGY--> [Dif obj (CMs: Text)]
      | --HAS_RIGHTS--> [Dif obj (CMs: Text, Still Image, Document)]
      | <--HAS_SUBMISSION_PACKET-- Dif obj (CMs: Email, Text, Still Image,
Document, Audio, Opaque)]
```

Relationship diagram showing what is possible in terms of relationships between files in a collection object, or the collection object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - No roles
- **Directory structure**
 - Because there aren't any content files, there aren't any directories.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.

Metadata Requirements

- Because there aren't any content files, this section does not apply.

Descriptor Details

- mets element
 - TYPE attribute = "SUBMISSION PACKET"
- In addition to the amdSec sections common to all content files:
 - May have a PREMIS event with metadata including the EAS event ID in the hulEventExtension schema
- No fileSec section
- A structMap section with a single empty div

CMID-12.0 (VIDEO) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** VIDEO
- **Content Model ID:** CMID-12.0

Summary Description

- **Description:** This content model represents moving images with potentially embedded audio, closed captions, subtitles and/or timecodes in digital form; could have been an email attachment.
- **Compatible object genres:** scanned film, television program, lecture, presentation, recorded event, concert

Related Software and Systems

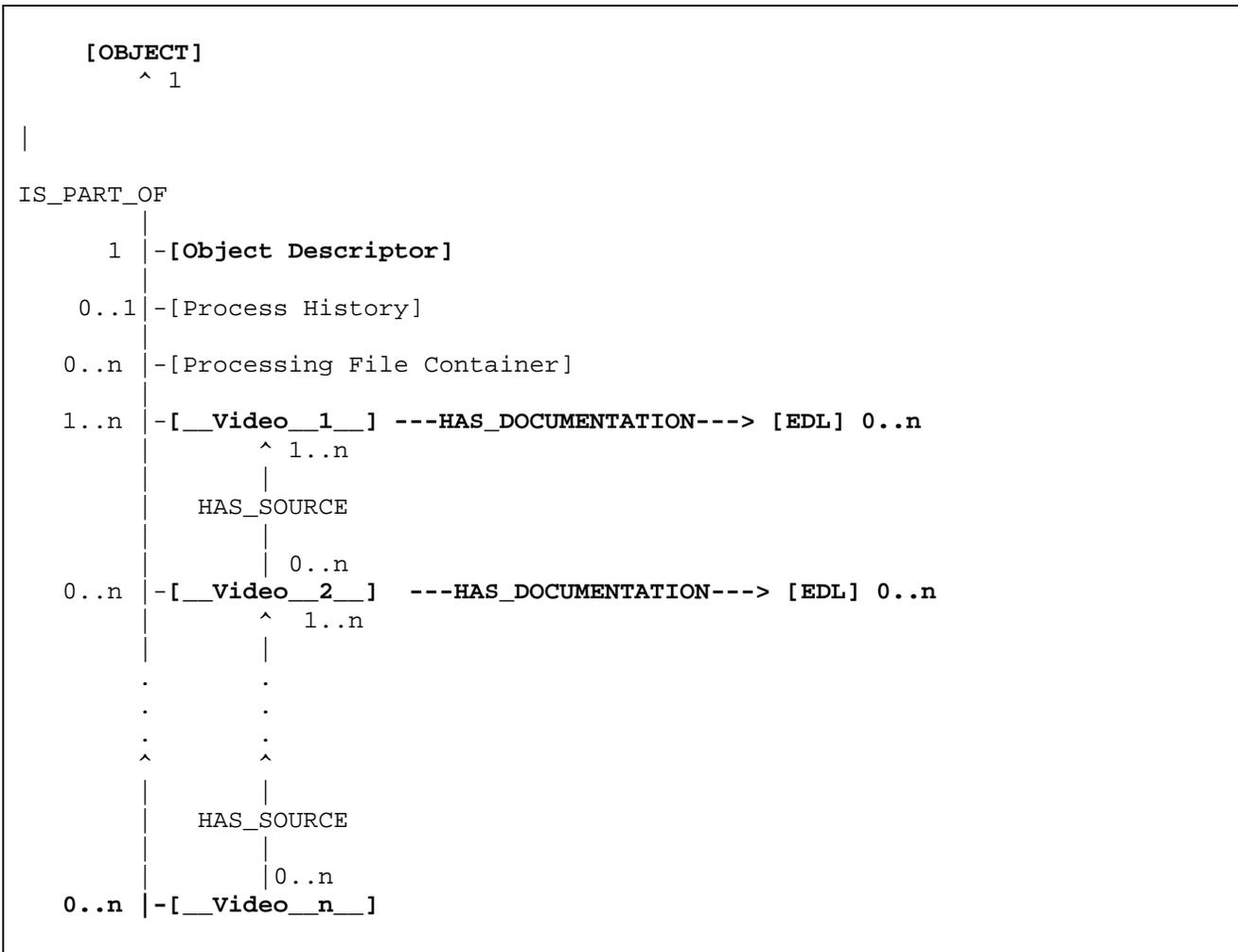
- **Delivery applications:** DRS Web Admin; SDS streaming delivery service
- **Rendering applications:** Web browsers

Repository Notes

- **Assessment and plans:** No plans or assessments but a few of the accepted but not preferred formats should be monitored closely (DNxHD, ProRes)

Content Model Details

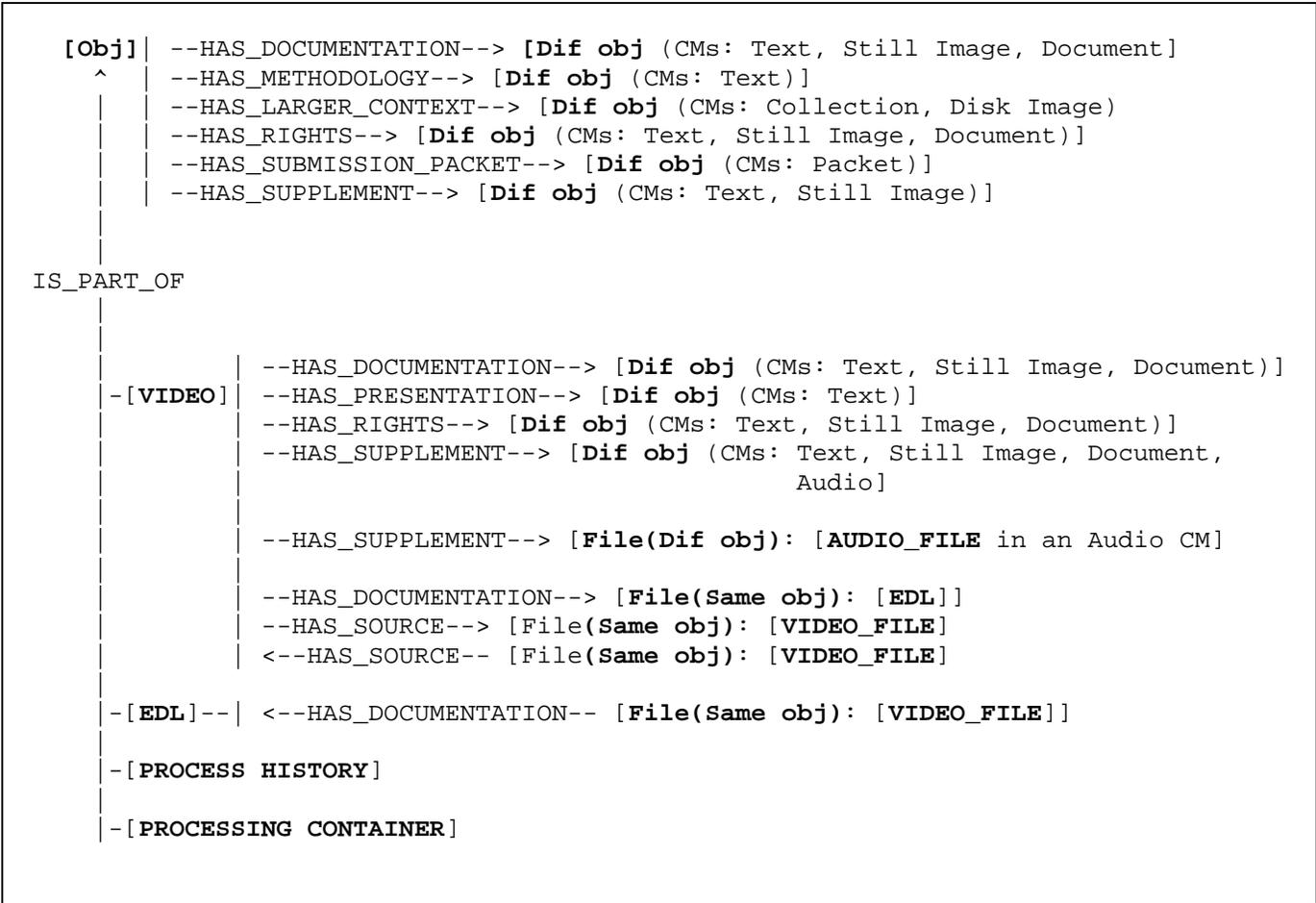
- **Structure:**



Structure diagram for a video object showing what is possible in terms of number of files and relationships within the object.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - Key relationships: The deliverable video files may be associated with presentation objects (containing AES60 XML playlist files) to provide delivery sequencing.
 - File-to-its Object
 - IS_PART_OF relationships between each child and the object
 - These are implicitly modeled by describing the files in the object descriptor.
 - File-to-File within the same Object
 - HAS_SOURCE relationships between derivative video files. **Note that multiple video files can be the source of another video file, for example when multiple archival masters were combined to form a production master.**
 - HAS_DOCUMENTATION relationship from each archival content file to each Edit Decision List. Note that as a business rule, all EDLs associated with a video file represent the same edit decisions, within the limits of their various EDL formats.
 - File-to-Different Object

- Any of the content files may be the source of HAS_RIGHTS or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
- The deliverable files may be the source of HAS_PRESENTATION relationships to objects containing AES60 playlists.
- A content file may be the source of a HAS_SUPPLEMENT relationship to another object (e.g. for closed caption data).
- Object-to-Different Object
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor
 - May be the source object of a HAS_LARGER_CONTEXT relationship between this object and its associated collection object (as in the case of email attachment objects) or to a disk image object
 - May be the source object of a HAS_SUBMISSION_PACKET relationship between this attachment object and its associated packet object.
 - May be the source of a HAS_SUPPLEMENT relationship to another object for a poster frame, or subtitle data
- File-to-File within Different Object
 - May be the source file for a HAS_SUPPLEMENT relationship to an audio description and/or double system audio file in another obje



Relationship diagram showing what is possible in terms of relationships between files in a video object, or the video object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models; File (Same obj) = File within the same object; File(Dif obj) = File within a different object)

- **Roles**
 - Object (may have one or more of the following roles)
 - EMAIL_ATTACHMENT
 - THESIS_SUPPLEMENT
 - Files
 - The video content files are differentiated by their role: ARCHIVAL_MASTER, PRODUCTION_MASTER or DELIVERABLE. Typically video files will have only one of these roles, but it is possible for them to have multiple roles, as in the case of archival masters that are in a format that can be delivered.
 - Process History files will have the role PROCESS_HISTORY.
 - Edit Decision List file(s) will have the role EDIT_DECISION_LIST
 - Processing File Container will have the role PROCESS_FILES
- **Directory structure**
 - Video files are contained in directories that begin with the prefix video

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.

- Suggested default: request URNs only for the content files having the role DELIVERABLE

Metadata Requirements

Object component	Metadata
Video content files	<ul style="list-style-type: none"> • Acceptable formats <ul style="list-style-type: none"> ○ 1 <ul style="list-style-type: none"> ▪ Codec format: JPEG 2000 ▪ Wrappers: QuickTime, MXF (MXF OP1a, OP1b operational patterns or AS-07) ▪ Notes: recommend lossless compression ○ 2 <ul style="list-style-type: none"> ▪ Codec format: Uncompressed ▪ Wrappers: QuickTime ▪ Notes: 8 bit or 10 bit ○ 3 <ul style="list-style-type: none"> ▪ Codec format: DV ▪ Wrappers: QuickTime ▪ Notes: For digitized DV tape ○ 4 <ul style="list-style-type: none"> ▪ Codec format: MPEG-2 ▪ Wrappers: QuickTime ▪ Notes ○ 5 <ul style="list-style-type: none"> ▪ Codec format: H.264 ▪ Wrappers: QuickTime ▪ Notes: Any of the 21 different profiles ○ 6 <ul style="list-style-type: none"> ▪ Codec format: Avid DNxHD ▪ Wrappers: QuickTime, MXF (MXF OP1a, OP1b operational pattern or AS-07) ▪ Notes ○ 7 <ul style="list-style-type: none"> ▪ Codec format: Apple ProRes ▪ Wrappers: QuickTime ▪ Notes • Metadata type: VIDEO <ul style="list-style-type: none"> ○ EBUCore technical metadata • Role: one or more of: <ul style="list-style-type: none"> ○ ARCHIVAL_MASTER <ul style="list-style-type: none"> ▪ Usage class: LOWUSE (suggested default) ▪ Access flag: N (suggested default) ○ PRODUCTION_MASTER <ul style="list-style-type: none"> ▪ Usage class: LOWUSE (suggested default) ▪ Access flag: N (suggested default) ○ DELIVERABLE <ul style="list-style-type: none"> ▪ Usage class: HIGHUSE (suggested default) ▪ Access flag: P (suggested default)

	<ul style="list-style-type: none"> ● One video in this object should have an isFirstGenerationInDRS value of true ● One video in this object should have an isPreferredDeliverableSource value of true ● Note that the audio codec will be Uncompressed PCM Audio but it is TBD if some formats (e.g. MPEG-2) will use a different audio codec
Process History file	<ul style="list-style-type: none"> ● Format: Extensible Markup Language <ul style="list-style-type: none"> ○ Process history is described by reVTMD. This file documents the complete history of the pre-DRS processing performed on this video object, including the digitization process for the master files, and the devices and settings used. ● MIME type: text/xml ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) Role: PROCESS_HISTORY
Edit Decision List file(s)	<ul style="list-style-type: none"> ● Format: Plain Text and/or Extensible Markup Language <ul style="list-style-type: none"> ○ Three types of EDLs are known candidates at this time: <ul style="list-style-type: none"> ■ CMX EDL, specification: http://xmil.biz/EDL-X/CMX3600.pdf ■ XMEML from Apple, specification: https://developer.apple.com/library/mac/documentation/AppleApplications/Reference/FinalCutPro_XML ■ Final Cut Pro X XML from Apple, specification: https://developer.apple.com/library/mac/documentation/FinalCutProX/Reference/FinalCutProXXMLFormat/Introduction/Introduction.html ● MIME type: text/plain or text/xml ● Metadata type: TEXT ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) Role: EDIT_DECISION_LIST
Processing Files Container	<ul style="list-style-type: none"> ● Format name: ZIP ● MIME type: application/zip ● Metadata type: none ● Usage class: LOWUSE (suggested default) ● Access flag: N (suggested default) ● Role: PROCESS_FILES (suggested default) <p>This ZIP file contains information about intermediate content files that were created as part of the processing but were deleted prior to deposit of the video content to the DRS. This intermediate file is included in this video object to provide information to the video processing lab about their pre-DRS processing. It is not expected that the DRS will know very much about these files. This ZIP file contains one or more files in the following formats:</p> <ul style="list-style-type: none"> ● 1. Format name: Extensible Markup Language

	<ul style="list-style-type: none"> ○ MIME type: text/xml ● 2. Format name: Edit Decision List <ul style="list-style-type: none"> ▪ See Edit Decision List section, above. ▪ MIME type: text/plain and/or text/xml ● 4. Format name: Unknown Binary ● MIME type: application/octet-stream
--	---

Descriptor Details

- mets element
 - TYPE attribute = “VIDEO”
- In addition to the amdSec sections common to all content files:
 - at most a single amdSec/digiprovMD section for process history file
 - mdRef@MDTYPE=”OTHER”
 - mdRef@OTHERMDTYPE = “reVTMD” [Note: this parallels the definition of audio descriptors, which treat the process history as a file rather than inline XML.]
 - at most one amdSec/sourceMD section containing object metadata for the source original.
 - mdWrap@MDTYPE=”OTHER”
 - mdWrap@OTHERMDTYPE = “UTVIDEOSRC”
- fileSec section (as specified for all content models)
- structMap section
 - TYPE attribute = “PHYSICAL”
 - Div sections with attribute LABEL one of archival, production, deliverable

CMID-14.0 (PDS DOCUMENT LIST) Status - IN-PRODUCTION

Definition Status

- **Completion status:** COMPLETED
- **Ingest support status:** IN-PRODUCTION

Identity

- **Aliases:** PDS DOCUMENT LIST
- **Content Model ID:** CMID-14.0

Summary Description

- **Description:** This content model represents a list of page-turned objects.
- **Compatible object genres:** series, multi-part works

Related Software and Systems

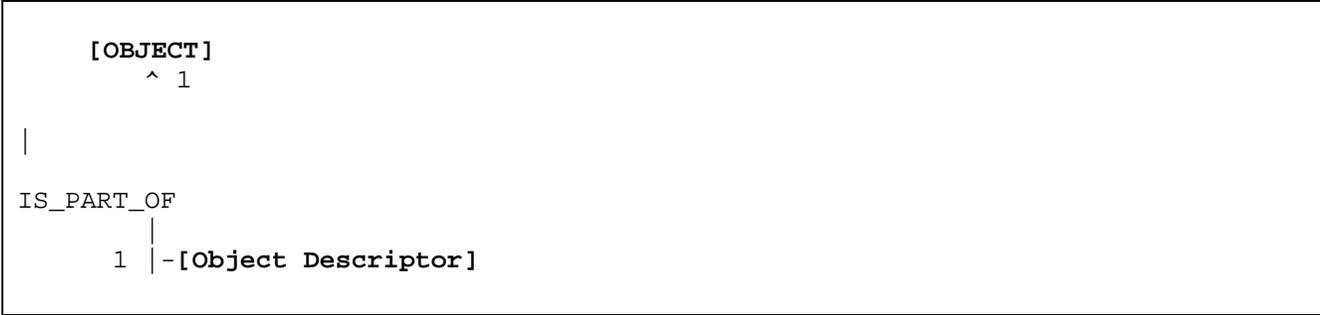
- **Delivery applications:** ListView, PDS, Havard Viewer
- **Rendering applications:** Web Browsers

Repository Notes

- **Assessment and plans:** No plans or assessments

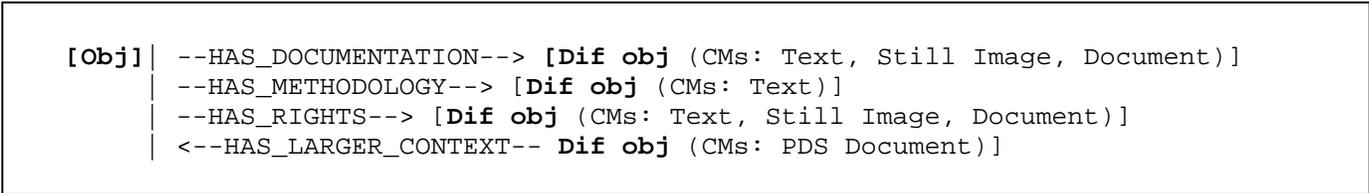
Content Model Details

- **Structure:**



Structure diagram for a list object showing what is possible in terms of number of files and relationships within the object.

- **Relationships** (unless otherwise noted these relationships are specified in the object descriptor using PREMIS relationships)
 - F-SO (File-to-Same Object)
 - N/A
 - F-F(SO) (File-to-File within Same Object)
 - N/A
 - F-DO (File-to-Different Object)
 - N/A
 - O-O (Object-to-Object)
 - May be the source object of HAS_RIGHTS, HAS_METHODODOLOGY or HAS_DOCUMENTATION relationships to another object.
 - HAS_RIGHTS relationships are expressed in PREMIS Rights metadata in the descriptor rather than as PREMIS relationships.
 - May be the target object of a HAS_LARGER_CONTEXT from objects in the PDS Document content model.
 - F-F(DO) (File-to-File within Different Object)
 - No relationships



Relationship diagram showing what is possible in terms of relationships between files in a collection object, or the collection object itself, and other files or objects. (Key: Dif obj = Different object; CMs = Content Models)

- **Roles**
 - Object
 - No roles
 - Files
 - No roles

- **Directory structure**
 - Because there aren't any content files, there aren't any directories.

Persistent Names

- Every object will automatically be assigned an NRS URN during the deposit process.

Metadata Requirements

- Because there aren't any content files, this section does not apply.

Descriptor Details

- mets element
 - TYPE attribute = "PDS DOCUMENT LIST"
- No fileSec section
- structMap section
 - TYPE attribute = "LOGICAL"
 - Div sections with attribute:
 - LABEL for the descriptive text to be used in user interfaces to identify the volume, e.g. Volume 2 (1888-1889)
 - ORDER for the sequence in which the volume is to be presented in the user interface, e.g. 2
 - ORDERLABEL for the range of page sequences contained in the volume relative of the entire extent of the material, e.g. 357-780