

# The new DRS (DRS 2) Content Model Reference

<b>Introduction: DRS objects .....</b>	<b>2</b>
<b>CMID-1.0 (OPAQUE) .....</b>	<b>4</b>
<b>CMID-1.1 (OPAQUE CONTAINER).....</b>	<b>6</b>
<b>CMID-2.0 (AUDIO) .....</b>	<b>9</b>
<b>CMID-4.0 (PDS DOCUMENT) .....</b>	<b>11</b>
<b>CMID-4.1 (DOCUMENT).....</b>	<b>13</b>
<b>CMID-5.0 (STILL IMAGE).....</b>	<b>15</b>
<b>CMID-5.2 (TARGET IMAGE).....</b>	<b>17</b>
<b>CMID-6.0 (TEXT).....</b>	<b>18</b>
<b>CMID-8.0 (COLOR PROFILE) .....</b>	<b>20</b>

October 16, 2013

© 2013 President and Fellows of Harvard College

<http://hul.harvard.edu/ois/systems/drs/docs/drs2-contentmodels.pdf>

## Introduction: 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 apply to all DRS content:

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

The following restrictions apply to PDS objects:

- PDS objects may contain up to 5,000 pages of content.
- Multiple PDS objects cannot be “merged” into one object for PDS display. Related functionality that will replace the need to “merge” objects and will allow display of multiple PDS objects in one PDS frame is being developed and will be available in a later release of the software.

### ***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 role OBJECT\_DESCRIPTOR.

These descriptor details are common to all content models:

- mets element
  - TYPE attribute indicates the content model:
    - OPAQUE
  - 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.

## CMID-1.0 (OPAQUE)

### 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 OIS 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

**Rendering applications:** none

### 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.

#### Accepted Formats

##### Content file:

- Any format  
Format name for unidentified formats: Unknown Binary
- MIME type: any  
MIME type for unidentified formats: application/octet-stream

##### Documentation file (any of the following):

- Plain Text
  - o MIME type: text/plain
  - o Character encoding: US-ASCII or UTF-8
  - o Technical metadata type: TEXT
- OpenDocument Text
  - o MIME type: application/vnd.oasis.opendocument.text
  - o Technical metadata type: DOCUMENT

- PDF/A-1a

Although this is preferred, Portable Document Format or any of the other PDF/A variations is acceptable

- o MIME type: application/pdf
- o Technical metadata type: DOCUMENT

- OpenDocument Spreadsheet

- o MIME type: application/vnd.oasis.opendocument.spreadsheet
- o Technical metadata type: N/A

**Roles****Object**

- May have an EMAIL\_ATTACHMENT role

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.
- No URNs requested for the content files

**Technical restrictions**

The following restrictions apply to all DRS content:

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

## CMID-1.1 (OPAQUE CONTAINER)

### 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 zipped. 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. video); any objects formatted differently than expected by OIS 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

### Content Model Details

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

#### Accepted formats

##### Compressed Archive:

- Format name: ZIP
  - o MIME type: application/zip
  - o Format technical metadata: none
  - o File metadata
    - Usage class: LOWUSE (suggested default)
    - Access flag: N (suggested default)

##### Content Files within the compressed archive (bitstreams):

- Any format
  - o format name for unidentified formats: Unknown Binary
  - o MIME type: any
  - o MIME type for unidentified formats: application/octet-stream

DRS documentation files within the compressed archive (bitstreams):

Any of the following:

- Format name: Plain Text
  - o MIME type: text/plain
  - o Character encoding: (not recorded for bitstreams)
  - o Technical metadata type: none
- Format name: OpenDocument Text
  - o MIME type: application/vnd.oasis.opendocument.text
  - o Technical metadata type: none
- Format name: PDF/A-1a
  - o although this is preferred, Portable Document Format or any of the other PDF/A variations are acceptable
  - o MIME type: application/pdf
  - o Technical metadata type: none
- Format name: OpenDocument Spreadsheet
  - o MIME type: application/vnd.oasis.opendocument.spreadsheet
  - o Technical metadata type: none

**Roles**

Object

- No roles

Files

- The compressed archive will have the role CONTAINER

**Directory structure**

The compressed archive will be located under a directory that begins with the prefix `container`.

- Within the compressed archive all the files will be contained in one of two top-level directories:
  - o `content`  
This directory contains content files (not documentation about the content).
  - o `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.
- No URNs requested for the compressed archive file.

### **Technical restrictions**

The following restrictions apply to all DRS content:

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



## CMID-2.0 (AUDIO)

### 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 for streaming (MP3, MP4/AAC formats only), FDS for download.

**Rendering applications:** many audio playback software options for MP3 and MP4/AAC including iTunes, VLC Media Player, Windows Media Player, Amorok, Banshee, xmms, and Rhythmbox.

### Content Model Details

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

#### Accepted Formats

##### Audio content files

- Audio Interchange File Format
  - MIME type: audio/x-aiff
- Broadcast Waveform Audio
  - MIME type: audio/x-wave
- MPEG 1/2 Audio Layer 3
  - MIME type: audio/mpeg
- MPEG Advanced Audio Encoding
  - MIME type: audio/mp4

##### Process History file

- Extensible Markup Language
  - MIME type: text/xml
  - Process history is described by AES-x098C: Administrative metadata for audio objects - Process history schema. 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.

##### Processing Files Container

- ZIP
  - MIME type: application/x-gzip
  - 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 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.

#### Audio Decision List

- Audio Decision List
  - MIME type: text/plain
  - 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

#### **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 synchronization.

#### **Roles**

- Object
  - May have an EMAIL\_ATTACHMENT role
- 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

#### **Persistent Names**

- Every object will automatically be assigned an NRS URN during the deposit process.
- Delivery URNs are requested only for the content files having the role DELIVERABLE

#### **Technical restrictions**

The following restrictions apply to all DRS content:

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

## CMID-4.0 (PDS DOCUMENT)

### Identity

**Aliases:** PDS DOCUMENT

**Content Model ID:** CMID-4.0

### Summary Description

**Description:** This content model represents a page-based object delivered by HUL'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)
- Image Delivery Service (IDS) for the page images
- Rendering applications: Web browser

### Content Model Details

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

### Accepted Formats

Page Image (per page image, one of the following:

- JPEG File Interchange Format
  - MIME type: image/jpeg
- Graphics Interchange Format
  - MIME type: image/gif
- Tagged Image File Format
  - MIME type: image/tiff
- JPEG 2000 JP2
  - MIME type: image/jp2

### Page Text

- UTF-8 encoded Plain Text
  - MIME type: text/plain

Page Text and Layout (XML files using the ALTO schema)

- Extensible Markup Language
  - MIME type: text/xml

**Roles****Object**

- No roles

**Files**

- All page images have the role PAGE\_IMAGE
- Page images may have one or more of the following roles: ARCHIVAL\_MASTER, PRODUCTION\_MASTER, DELIVERABLE
- All page text files have the role PAGE\_TEXT
- All page layout files have the role PAGE\_COORDINATES

**Directory structure**

- 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.
- Every object will be automatically assigned a PDS delivery URN.
- IDS delivery URNs may be requested pre-deposit and post-deposit for the content page image files having the role DELIVERABLE
- FDS delivery URNs may be requested post-deposit for all text files.

**Technical restrictions**

The following restrictions apply to all DRS content:

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

The following restrictions apply to PDS objects:

- PDS objects may contain up to 5,000 pages of content.
- Multiple PDS objects cannot be “merged” into one object for PDS display. Related functionality that will replace the need to “merge” objects and will allow display of multiple PDS objects in one PDS frame is being developed and will be available in a later release of the software.

## CMID-4.1 (DOCUMENT)

### 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 HUL'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

### Content Model Details

**Structure:** one or more document files, potentially having derivative document files.

**Accepted Formats (one of the following):**

- Portable Document Format (PDF)
- PDF/A
- PDF/X

### Roles

Object (may have one of these roles):

- DOCUMENTATION
- DONOR\_AGREEMENT
- EMAIL\_ATTACHMENT
- FINDING\_AID
- HARVARD\_POLICY
- LICENSE
- RISK\_ASSESSMENT
- STATUTE

Files (may have one or more 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.

A Delivery URN will be requested for each of this object's content files.

**Technical restrictions**

The following restrictions apply to all DRS content:

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

## CMID-5.0 (STILL IMAGE)

### 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)

### Content Model Details

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

#### Accepted Formats:

Image (per image, one of the following):

- JPEG File Interchange Format
  - MIME type: image/jpeg
- Graphics Interchange Format
  - MIME type: image/gif
- Tagged Image File Format
  - MIME type: image/tiff
- JPEG 2000 JP2
  - MIME type: image/jp2

#### Roles

Object (may have one of the following roles):

- DOCUMENTATION
- DONOR\_AGREEMENT
- FINDING\_AID
- HARVARD\_POLICY
- LICENSE
- RISK\_ASSESSMENT

- STATUTE

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**

Images are contained in directories that begin with the prefix “image” and any subdirectories are also named with the prefix “image”.

#### **Persistent Names**

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

Delivery URNs will be requested only for the content files having the role DELIVERABLE.

#### **Technical restrictions**

The following restrictions apply to all DRS content:

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



## CMID-5.2 (TARGET IMAGE)

### 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)

### Accepted Formats:

#### Target Image:

- TIFF (MIME type: image/tiff)
- JPEG (MIME type: image/jpeg)
- JP2 (MIME type: image/jp2)
- GIF (MIME type: image/gif)

#### Target Description:

- Plain Text (MIME type: text/plain)

### Roles

#### Object

- No roles

#### Files

- TARGET\_IMAGE role for the target image
- TARGET\_DESCRIPTION role for the target description file

### Directory structure

Target images are contained in a directory that begins with the prefix “image”

Target description files are contained in a directory that begins with the prefix “tdf”

### Persistent Names

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

No URNS are requested for the content files

## CMID-6.0 (TEXT)

### Identity

**Aliases:** TEXT

**Content Model ID:** CMID-6.0

### Summary Description

**Description:** This content model represents relatively simple text objects. 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, Audio Decision Lists, Email attachments, audio playlists.

**Audio Decision Lists - According to the AES website:**

“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.”

### Related Software and Systems

**Delivery applications:** DRS Web Admin, File Delivery Service (FDS), Streaming Delivery Service (SDS) for audio playlists.

**Rendering applications:** Web browser, general-purpose text viewer or editor (WordPad, TextPad, vi, emacs, etc.), or format-specific (XML, SGML) editor

### Content Model Details

**Structure:** one or more text files with optional derivative files. In the future text components may include additional files associated with text files (schema, etc.).

#### Accepted Formats:

##### Text file:

- Standard Generalized Markup Language (MIME type: text/sgml)
- Extensible Markup Language (MIME type text/xml)
- Plain Text (MIME type: text/plain)
- Audio Decision List (MIME type: ?)

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)

### Roles

Object (may have one of the following):

- CODE

- DOCUMENTATION
- DONOR\_AGREEMENT
- FINDING\_AID
- HARVARD\_POLICY
- LICENSE
- PRESENTATION
- RISK\_ASSESSMENT
- SCHEMA
- STATUTE
- 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.

A Delivery URN will be requested for each of this object's content files.

## CMID-8.0 (COLOR PROFILE)

### 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>).

### Content Model Details

**Structure:** a single color profile.

Accepted Formats:

ICC File Format for Color Profiles (MIME type: application/x-icc)

### Roles

Object (may have one of the following)

- COLOR\_PROFILE

### Directory structure

The color profile file is located within a directory having the prefix “color\_profile”

### Persistent Names

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

No delivery URNs are requested for the color profile files.