

# FADGI Guidelines for Embedded Metadata within DPX File Headers

Kate Murray  
Library of Congress

IASA Annual Meeting  
Berlin Germany 2017

# FADGI: Federal Agencies Digital Guidelines Initiative

[Glossary](#) | [Participants](#) | [Contact Us](#)

Search this site

**FA DGI** Federal Agencies Digital Guidelines Initiative

[Submit comments](#) on guidelines or on our digitization efforts in general or [Tweet #FADGI](#)

[Home](#) [About](#) [Guidelines](#) [Resources](#) [Review Process](#) [News & Events](#)

[Print](#) [Subscribe](#) [Share/Save](#)

## About This Initiative

FADGI is a collaborative effort started in 2007 by federal agencies to articulate common sustainable practices and guidelines for digitized and born digital historical, archival and cultural content. Two working groups study issues specific to two major areas, Still Image and Audio-Visual.

**Summary chart** of FADGI's impact within the Library of Congress and the wider community. Updated September 8, 2017.

[Learn more about the initiative »](#)

### Still Image Working Group

This group is involved in a cooperative effort to develop common digitization guidelines for still image materials.

### Audio-Visual Working Group

This group works collaboratively on common and sustainable technical guidelines, methods, and practices for digitized and born

## FADGI Guidelines

### **MXF AS-07: Revised Specification and New Sample Files Published**

Approved by Working Group | September 8, 2017

A revised version of the MXF AS-07 Application Specification, along with a new set of corresponding graded sample files, is now available. AS-07 is developed by the FADGI Audio-Visual Working Group in conjunction with community and commercial collaborators and specifies a vendor-neutral subset of the MXF file format for the long-term archiving and preservation of moving image and other audiovisual content. Among other features, AS-07 defines a means for the carriage and labeling of multiple timecodes and audio tracks; the handling of captions, subtitles, and Timed Text; a minimal core metadata set; program segmentation metadata; and embedded content integrity data.

### **Guidelines for Embedding Metadata in DPX File Headers**

Approved by Working Group | August 14, 2017

The FADGI Audio-Visual Working Group has developed guidelines for embedding metadata in the DPX header, a file format often used for digitized motion picture film. The guidelines outline FADGI implementations of the SMPTE Core fields as well as other elements *Strongly Recommended*, *Recommended* or *Optional* for FADGI use. The approved version includes significant revisions from the 12/16/16 draft version such as a justification about the rationale for embedded metadata and explanations of issues with date/time formatting and data overruns.

## News & Events

- Kate Murray will introduce the [DPX Embedded Metadata Guidelines](#) at the [IASA Annual Conference](#) in September 2017
- Putting the "Digital" in the [FADGI acronym](#), January 2017
- FADGI documents to include [CC0 1.0 Universal License](#), December 2016
- [Poster about DPX embedded metadata project](#) at 2016 AMIA conference, November 2016
- Tom Rieger talks digital strategy at Internet Archives' [Library Leaders Forum](#), October 2016
- FADGI's Film Scanning SOW highlighted on [Perfs to DPX and All Points In Between: Perspectives on Current Film Digitization Practices](#) panel discussion at IASA 2016 Annual Conference, September 2016

[See more News & Events »](#)

## Blog Posts

- [FADGI's 10th Anniversary: Adapting to Meet the Community's Needs](#), February 6, 2017
- [New FADGI Guidelines for Embedded Metadata in DPX Files](#), December 16, 2016
- [FADGI MXF Video Specification Moves Up an Industry-organization Approval Ladder](#), July 11, 2016

[See more Blog Posts »](#)

<http://www.digitizationguidelines.gov>

# The FADGI Embedded Metadata Family

## **Embedding Metadata in Digital Audio Files Guideline for Federal Agency Use of Broadcast WAVE Files**

By the Federal Agencies Audio-Visual Working Group  
<http://www.digitizationguidelines.gov/audio-visual/>

Version 2. Approved version, April 23 2012.

Home / Browse / BWF MetaEdit



## BWF MetaEdit

Brought to you by: [fadgi](#), [I0c](#)

Summary | Files | Reviews | Support | Wiki | Tickets ▾ | Discussion | Code

★ 4.8 Stars (15)  
↓ 86 Downloads (This Week)  
📅 Last Update: 2014-04-25

[Tweet](#) [G+](#) [Like 8](#)

 [Download](#)  
BWF\_MetaEdit\_GUI\_1.3.1\_Mac.dmg

[Browse All Files](#)



**Federal Agencies  
Digital Guidelines Initiative**

**AS-07: MXF Archive and Preservation Format  
Application Specification**

New version and  
sample files released  
in early September  
2017

- 6.6 Descriptive Metadata Parameters and Constraints .....
- 6.6.1 AS-07 Descriptive Metadata (informative).....
- 6.6.1.1 AS-07 Core Descriptive Metadata Scheme (informative).....
- 6.6.1.1.1 AS-07 Core DMS Device Objects (informative).....
- 6.6.1.2 AS-07 DMS Identifier Objects (informative).....
- 6.6.1.3 AS-07 Generic Stream Partition Superclass Descriptive Metadata Scheme (informative).....
- 6.6.1.3.1 AS-07 Generic Stream Partition Binary Data Descriptive Metadata Framework (informative) .....
- 6.6.1.3.2 AS-07 Generic Stream Partition Text-based Data Descriptive Metadata Framework (informative) .....
- 6.6.1.4 AS-07 Segmentation Descriptive Metadata Scheme (informative) .....
- 6.6.1.4.1 AS-07 Segmentation Descriptive Metadata Scheme – Parts Object (informative) .....
- 6.6.2 AS-07 Descriptive Metadata Schemes Encoder Requirements .....
- 6.6.2.1 AS-07 Descriptive Metadata Track Encoder Requirements.....
- 6.6.2.2 AS-07 Descriptive Metadata Track Decoder Requirements .....
- 6.6.3 Shim Parameter Table for Descriptive Metadata Schemes.....
- 6.6.4 Redundant Metadata.....
- 6.6.5 KLV Fill.....
- 6.6.6 Static Descriptive Metadata Requirements .....

# TIFF Headers

## **Guidelines: Embedded Metadata in TIFF Images**

Still Image Working Group

*Embedded metadata can travel with a digital object during its life cycle and often exists in synergy with metadata in an organization's databases or other information technology systems. Embedded metadata enables people in and outside an organization to work more efficiently, provides valuable data to the systems that preserve digital content, and can assist in disaster recovery. In addition to these guidelines for minimal descriptive metadata in still image files, FADGI offers embedding guidelines for minimal descriptive metadata in still images and for Broadcast WAVE files in audio-visual materials.*

### **Document Background**

The draft guideline on TIFF Image Metadata submitted by the Still Image Working group represents a minimal set of recommended embedded metadata for historical and cultural heritage digital imaging. This is presented as an interim guideline. A new recommended guideline encompassing a richer metadata set in a more flexible format, with wider file format support, is expected to be developed by the Embedded Metadata Sub-group and posted at a future date.

# Typical FADGI Metadata: High Level

Provide enough context to understand what the file is and where to get more info about it:

- Identifiers
- Ownership statement/contact info
- File creation date
- Copyright & use status (if known)
- Other:
  - Info about structure & history of the file
  - Fixity info

# Why DPX?

- DPX highly adopted in US federal community
- Companion to FADGI's [Digitizing Motion Picture Film: Exploration of the Issues and Sample SOW](#) (April 2016)
- Initiated by National Museum of African American History and Culture in **Spring 2016**
- Existing tool not meeting their needs
- [AMIA Hack Day 2015](#)
- First come the guidelines, then comes the tool



# Spec Says One Thing, Files Say Another

- Analyzed conformance of FADGI member files to SMPTE ST268
- The sample files were created by film scanners from a range of well-known manufactures and contain a variety of imagery content, resolutions, bit depths and more.
- Implementation of core and non-core fields are not always consistent according to the SMPTE ST268 specification set.
- Anonymized sample files analysis:  
<http://www.digitizationguidelines.gov/guidelines/digitize-DPXembedding.html>



# CORE fields

- To provide a streamlined path for implementation and testing, SMPTE defines a core set of required fields which contains the minimum amount of information that a DPX reader needs to read and interpret a file.
- According to SMPTE ST268, a core-compliant reader must read the core fields, but need not read the others; and a core-compliant writer must fill the core fields with valid values

# CORE field issues

<b>Field number</b>	<b>Content</b>	<b>Summary of Analysis</b>
2	Offset to image data in bytes	Inconsistent. Sometimes used correctly, sometimes this data is recorded in field 19 or 21.12  *See 7.1.3 for more information on this field's relationship to field 77, Image Data Block Boundary
4	Total image file size in bytes (including file header)	Inconsistent. Sometimes used correctly, sometimes blank
21.12	Offset to image data	Inconsistent. Sometimes used correctly, sometimes this data is recorded in field 2 or 19

**Table 1. SMPTE Core Field Issues**

# Non-CORE Field Issues

Field number	Content	Summary of Analysis
10	Creation date/time: <u>yyyy:mm:dd:hh:mm:ssLTZ</u>	Inconsistent. Sometimes used correctly but often blank or the formatting is incorrect with dashes instead of colons or data in the wrong order. This value is hard to determine in post-production so embedding at time of creation would be the best course of action.
49	Format - e.g. Academy	Inconsistent. Sometimes used as summary for source format
56	Slate information	Inconsistent. Sometimes used correctly, sometimes used as summary of source format
66	Temporal Sampling or frame rate in HZ	Inconsistent. Sometimes replicates data in field 53 sometimes blank.

# Other Issues/ST 268 Errors

- SMPTE ST 268 does not provide definitions for all fields so it's a best guess sometimes
- Magic Number Data Type Mismatch should be labeled ASCII, not U32
- Image data offset mismatches
- Data field overruns
- Incorrect use of reserved fields



# **Federal Agencies Digital Guidelines Initiative**

## **Guidelines for Embedded Metadata within DPX File Headers for Digitized Motion Picture Film**



This work is available for worldwide use and reuse under [CC0 1.0 Universal](https://creativecommons.org/licenses/by/4.0/).

# FADGI Guidelines

- Outline FADGI implementations of the SMPTE Core fields as well as other elements *Strongly Recommended, Recommended or Optional*
- The non-Core fields take advantage of existing header structures as well as define new metadata elements for the User Defined fields to document, among other things, digitization process history

# FADGI Guidelines

- No obligation for non-CORE fields not of interest to FADGI
- Provide definitions for all fields of interest
- FADGI concurs with SMPTE defined use of CORE fields
- FADGI defined use for non-CORE fields, including User Defined Fields



# ST 268 Correction

<b>Magic number</b>	
SMPTE ST 268 Definition	Indicates the start of the image file and is used to determine byte order. The file format allows machines to create files in either of the two most common byte orders, whichever is easier for that machine. Byte-order translation is only required for machines reading files that were created on a machine with reverse byte order. Programs creating DPX files should write the magic number with the ASCII value of "SDPX" (0x53445058 hex). Programs reading DPX files should use the first four bytes to determine the byte order of the file. The first four bytes will be S, D, P, X if the byte order is most significant byte first, or X, P, D, S if the byte order is least significant byte first.
FADGI Use	Same as SMPTE ST 268
Field number	1
Byte offset start	0
Byte offset end	3
Byte limit	4
Data type	ASCII (ST 268 declares that the Magic Number for DPX is U32 data type but this incorrect. The data type should be labelled as ASCII.)
Header location	File information header
Mandatory/optional	Required by SMPTE ST 268
Values	SDPX; XDPX

# FADGI Provided Definition

## Version number of header format (V1.0 or V2.0)

SMPTE ST 268  
Definition

*[ST268 has no formal definition for this field]*

FADGI Use

Same intended use as SMPTE ST 268. Declaration of the version number of the DPX format. There are currently two versions of DPX: Version 1 (V1.0) defined by ST 268:1998 and Version 2 (V2.0) defined by ST 268:2003 and Amd 1.

Field number

3

Byte offset start

8

Byte offset end

15

Byte limit

8

Data type

ASCII

Header location

File information header

Mandatory/optional

Required by SMPTE ST 268

Values permitted by  
ST 268

V1.0  
V2.0

# Digitization Process History

Values	<p>The first line documents the source film reel, the second line contains data on the capture process and the third line contains data on the storage of the file. A new line is added when the coding history related to the file is changed.</p> <p>Each variable within a string is separated by a comma-space and each line should end with a carriage return and line feed. Each variable is optional, to be used when needed.</p> <p>O=format (reversal, print, positive, negative, DPXv1, DPXv2, etc.) G=gauge (super8mm, 8mm,16mm, 35mm, etc.) C=color (color, BW) S=sound (silent, composite optical, composite mag, separate optical reel, separate mag reel, etc.) D=summary of condition issues, especially if condition impacts visual quality of digitized image F=frames per second A=aspect ratio L=timing, grading (one-light, scene) W=bit depth (12-bit, 10-bit, 8-bit, etc.) R=resolution (2K, 4K, 8K, etc.) M=color model (RGB Log, etc.) N=name of vendor or operator who scanned film (if applicable) T=free ASCII text string; contains no commas but semicolons may be used.</p>
--------	--

# Example

O=positive, G=16mm, C=color, S=silent, F=24, A=4:3,D=warped  
O=DPXv1, L=one-light, W=10-bit, R=2K, M=RGB Log,  
T=FilmScannerA; SN123456; in-house  
O=DPXv1, W=10-bit, R=2K, M=RGB Log

[Explanation:

Line 1 reads: a 16mm positive color print, with no associated soundtrack, at 24fps and 4:3 aspect ratio (1.375:1). The film was warped and impacted the visual quality of the image.

Line 2 reads: film was digitized to a DPX version 1 file. One-light grading was employed. The image is 10-bit at 2K resolution (2048x1556) with RGB Log color model. The film was digitized via FilmScannerA (anonymized name of film scanner), serial number 123456, which is an in-house film scanner.

Line 3 reads: the file is stored as DPXv1, 10-bit 2K RGB log]

Example

# What's next?

- Involvement with SMPTE TC-31FS DG  
Constrained DPX for HDR
- Planning for DPX MetaEdit, open source batch  
embedding tool with limited QC operations  
(fingers crossed for FY18)

# Thank you!

Kate Murray

[kmur@loc.gov](mailto:kmur@loc.gov)

@fileformatology

<http://www.digitizationguidelines.gov/guidelines/digitize-DPXembedding.html>