



# Transfer Guidance Revision Project: Identifying “Fit-for-Purpose” Formats



Donald Chalfant  
Kevin DeVorse  
Louise Guenther  
Kathryn Nielsen

# Presentation Overview

- What We've Learned: Ten years with the existing transfer guidance for federal agencies
- What are our goals?
- Technical White Papers.
- A new look.
- Metadata for transfer.
- Q&A

# Current Guidance (developed between 2002-2004):

THE U.S. NATIONAL ARCHIVES & RECORDS ADMINISTRATION

[www.archives.gov](http://www.archives.gov)

Printed on: Friday, October 22, 2010

[Code of Federal Regulations]

[Title 36, Volume 3]

[Revised as of July 1, 2003]

From the U.S. Government Printing Office via GPO Access

[CITE: 36CFR1228.270]

[Page 725-728]

TITLE 36--PARKS, FORESTS, AND PUBLIC PROPERTY

CHAPTER XII--NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

PART 1228--DISPOSITION OF FEDERAL RECORDS--Table of Contents

Subpart L--Transfer of Records to the National Archives of the United States

## Sec. 1228.270 Electronic records.

(a) Timing of transfers. Each agency is responsible for the integrity of the permanent records it transfers on physical media to the National Archives of the United States. For records transferred by a media-less method, NARA works with the agency to ensure integrity of the records during the transfer process. To ensure that permanent electronic records are preserved, each Federal agency must transfer electronic records to NARA promptly in accordance with the agency's records disposition schedule. Furthermore, if the agency cannot provide proper care and handling of the media (see part 1234 of this chapter), or if the media are becoming obsolete and the agency cannot migrate the records to newer media, the agency must contact NARA to arrange for timely transfer of permanent electronic records, even when sooner than provided in the records schedule.

## Reflect NARA's capabilities at the time.

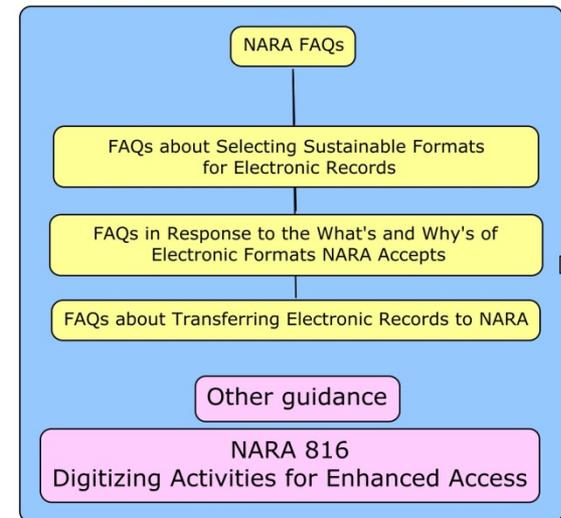
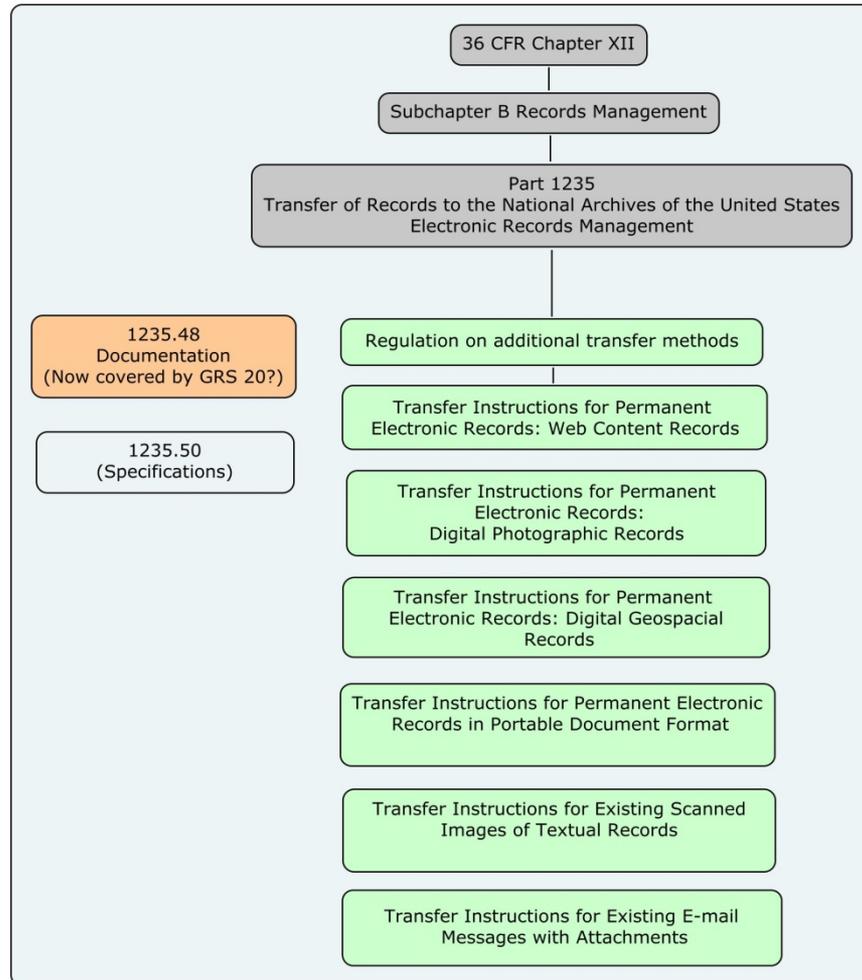
## FAQs in Response to the What's & Why's of Electronic Formats NARA Accepts

### *What formats of electronic records does NARA accept?*

According to its regulations (36 CFR 1228.270(d)) and transfer guidance developed during the Electronic Records Management [ERM] E-Government Initiative, NARA accepts over twenty formats (or versions thereof) of electronic records, ranging from digital photography to web pages to geographic information systems. In addition to the formats cited in the regulations, additional formats that NARA accepts can be found at <http://www.archives.gov/records-mgmt/initiatives/erm-products.html>.

NARA is currently developing an Electronic Records Archives [ERA] system that will allow it to accept even more formats of electronic records in the future.

# Current Guidance: limited scope that does not address all record/format types.



## Issues with current guidance products:

- Demonstrate a preference for standards based formats that agencies do not always use.
- Require that agencies transform to acceptable formats prior to transfer.
- Have proven an obstacle to NARA and agencies.

# Goals for revised guidance:

- Provide clear, concise, and consistent direction to agencies regarding formats that are acceptable for use when transferring records to NARA.
- Develop a flexible and extensible framework that can adapt to future needs.
- Balance preference for open formats with the business needs of agencies.



# Goals for revised guidance cont.:

- Match record behavior and performance to the correct file formats.
- Adapt to the change from tapes to ERA.
- Support a life-cycle approach from creation through to researcher access.

# Goals for revised guidance cont.:

- Expand the types of formats that NARA accepts.
- Acknowledge formats that are ubiquitous in the market place.
- Minimize the need for agencies to transform records prior to transfer.

# Where are we now?

- Revision project is almost ready for internal review.
- We've identified the categories and file formats that agencies are using today.
- We've conducted analysis of about 50 formats.
- We are compiling the results in the form of a revised guidance product.
- We are developing minimal metadata guidance.

# A Change in Approach

- We will identify formats that are:
  - Preferred
  - Acceptable
  - Acceptable for Imminent Transfer (to sunset previously acceptable formats)
  - Not Acceptable (this will be rare)

# Categories of E-records

- Digital Still Images
- Digital Moving Images
- Digital Audio
- Text
- Geospatial Records
- CAD
- Structured Data
- E-mail
- Web & Social Media

# Measuring sustainability\*:

- **Disclosure:** the degree to which complete specifications and technical integrity tools exist.
- **Adoption:** the degree to which the format is used by creators, disseminators, or users.
- **Transparency:** the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor.
- **Self-documentation:** formats that contain all the metadata needed to render the data as usable information.
- **External dependencies:** refers to the degree to which a format depends on particular hardware, operating system, or software for rendering or use.
- **Impact of patents:** Patents related to a digital format may inhibit the ability of archival institutions to sustain content in that format.
- **Technical protection mechanisms:** To preserve digital content and provide service to users and designated communities decades hence, NARA must be able to replicate the content on new media, migrate and normalize it in the face of changing technology, and disseminate it to researchers.

\*adapted from <http://www.digitalpreservation.gov/formats/>

# Measuring Sustainability.

- Analyzing format resources.
  - ISO/ANSI and other specifications
  - Corporate sites
  - LC Format Sustainability of Digital Formats
  - TNA-UK PRONOM Format Registry
  - CDL/LC UDFR
  - Wikipedia

# Determining “Fit for Purpose” Formats.

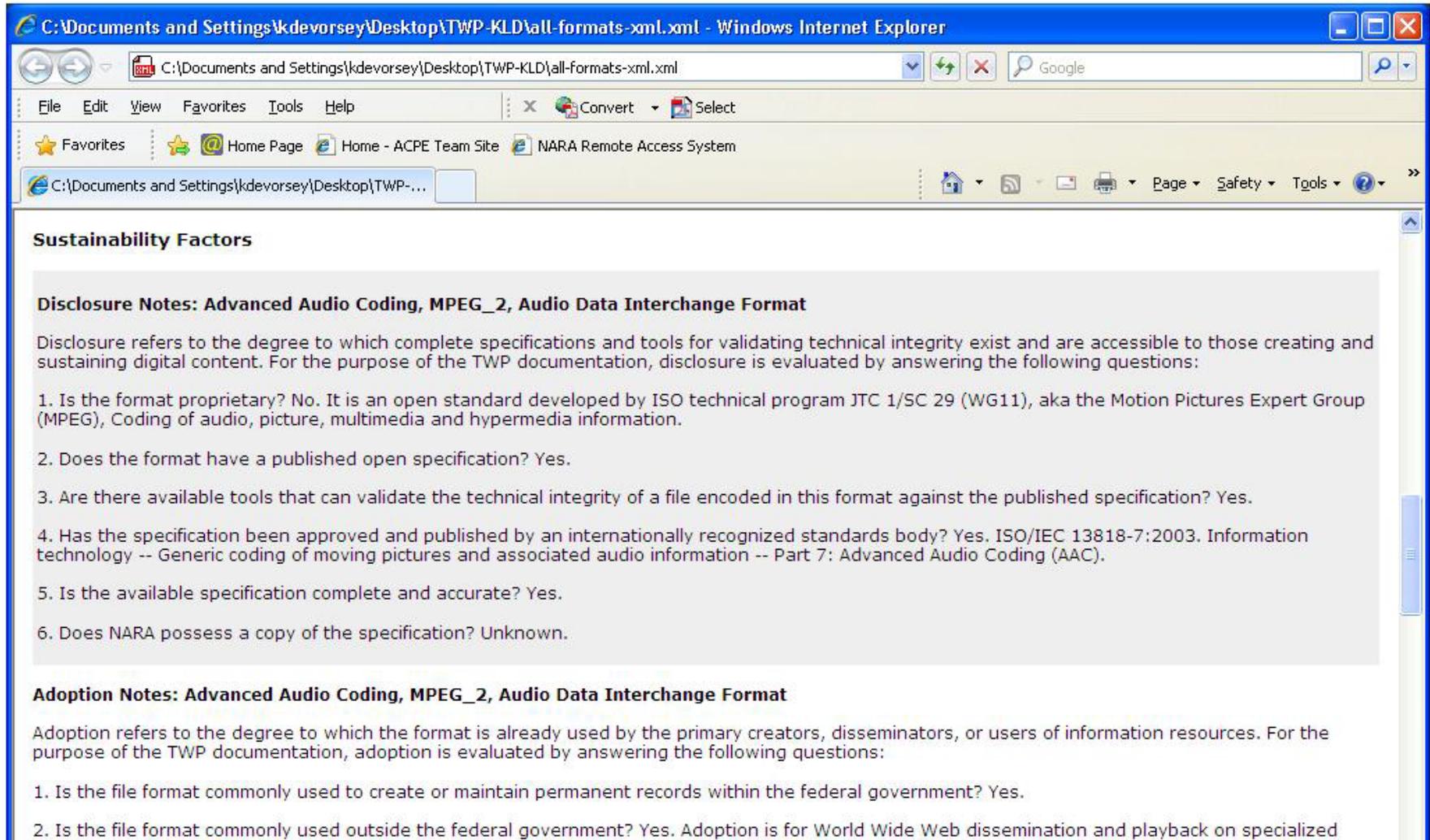


- The conventional wisdom is not always correct.
- Open formats may not always be the best choice.
- Proprietary formats aren't always so bad.

## GIF87A: Graphics Interchange Format [Version=87a]

### Properties

<b>Format Name</b>	Graphics Interchange Format
<b>Scope</b>	
<b>Alternate Name</b>	<ul style="list-style-type: none"> <li>• Graphics Interchange Format, Version 87a</li> <li>• GIF_87a</li> </ul>
<b>Version Number</b>	87a
<b>Digital Record Type</b>	format
<b>Specification Type</b>	Fully-documented-but-proprietary-format
<b>Orientation</b>	binary
<b>PRONOM Id</b>	fmt/3
<b>LOC Id</b>	
<b>Other Id</b>	Apple Uniform Type Identifier: com.compuserve.gif
<b>External Signatures</b>	<ul style="list-style-type: none"> <li>• gif</li> <li>•</li> </ul>
<b>File Extensions</b>	<ul style="list-style-type: none"> <li>• gif</li> <li>•</li> </ul>
<b>Mime Types</b>	image/gif
<b>Internal Signatures</b>	<p>Located in the header</p> <p>byte-offset: 0</p> <p>position-type: absolute-from-bof</p>
<b>Magic Numbers</b>	<ul style="list-style-type: none"> <li>• Little-endian (Intel) Hex: 47 49 46 38 37 61</li> <li>• Little-endian ASCII: G I F 8 7 a</li> </ul>
<b>Byte Order</b>	Big Endian
<b>Technical Properties</b>	<p>GIF87a is a file format for defining generalized color raster image data. GIF87a supports images with up to 8 bits per pixel (or 256 colors) on a variety of graphics hardware.</p> <p>GIF files are organized into seven sections: the GIF Signature, the Screen Descriptor, the Global Color Map, the Image Descriptor, the Local Color Map, Raster Data, and the GIF Terminator. Only the GIF Signature, the Screen Descriptor, the Image Descriptor, Raster Data, and the GIF Terminator are required while the Global and Local Color Maps are optional.</p> <p>The GIF format supports animation by allowing for the inclusion of multiple images in a single file. Each image descriptor is "introduced" by the Image Separator that is the hex characters 0x2C or ',' (comma).</p> <p>Components: The GIF Signature is the internal identifier Hex: 47 49 46 38 37 61 or the ASCII characters G I F 8 7 a</p> <p>The Screen Descriptor describes the overall parameters for all GIF images that are contained in the file including the dimensions for the image space, the existence of color mapping, the background screen color, and color depth information all described in a series of 8 bit bytes.</p>



C:\Documents and Settings\kdevorse\Deskto\TWP-KLD\all-formats-xml.xml - Windows Internet Explorer

C:\Documents and Settings\kdevorse\Deskto\TWP-KLD\all-formats-xml.xml

File Edit View Favorites Tools Help

Convert Select

Home Page Home - ACPE Team Site NARA Remote Access System

C:\Documents and Settings\kdevorse\Deskto\TWP-...

## Sustainability Factors

### Disclosure Notes: Advanced Audio Coding, MPEG\_2, Audio Data Interchange Format

Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. For the purpose of the TWP documentation, disclosure is evaluated by answering the following questions:

1. Is the format proprietary? No. It is an open standard developed by ISO technical program JTC 1/SC 29 (WG11), aka the Motion Pictures Expert Group (MPEG), Coding of audio, picture, multimedia and hypermedia information.
2. Does the format have a published open specification? Yes.
3. Are there available tools that can validate the technical integrity of a file encoded in this format against the published specification? Yes.
4. Has the specification been approved and published by an internationally recognized standards body? Yes. ISO/IEC 13818-7:2003. Information technology -- Generic coding of moving pictures and associated audio information -- Part 7: Advanced Audio Coding (AAC).
5. Is the available specification complete and accurate? Yes.
6. Does NARA possess a copy of the specification? Unknown.

### Adoption Notes: Advanced Audio Coding, MPEG\_2, Audio Data Interchange Format

Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information resources. For the purpose of the TWP documentation, adoption is evaluated by answering the following questions:

1. Is the file format commonly used to create or maintain permanent records within the federal government? Yes.
2. Is the file format commonly used outside the federal government? Yes. Adoption is for World Wide Web dissemination and playback on specialized

	Preferred	Acceptable	Imminent Transfer
<b>Structured Data</b>	<ul style="list-style-type: none"> <li>• ASCII Text</li> <li>• CSV, Comma Separated Values (RFC 4180)</li> <li>• XML, Extensible Markup Language</li> </ul>	<ul style="list-style-type: none"> <li>• DBF, dBASE Table File Format</li> <li>• HDF5, Hierarchical Data Format Version 5</li> <li>• CDF, Common Data Format</li> </ul>	<ul style="list-style-type: none"> <li>• EBCDIC</li> </ul>
<b>Digital Audio</b>	<ul style="list-style-type: none"> <li>• WAVE_LPCM_BWF, Broadcast WAVE Audio File Format</li> <li>• WAVE, WAVE Audio File Format"</li> </ul>	<ul style="list-style-type: none"> <li>• FLAC_1_1_2, FLAC (Free Lossless Audio Codec), Version 1.1.2</li> <li>• Ogg, Ogg File Format</li> <li>• MP3_ENC, MP3 Audio Encoding (MPEG Layer III Audio Encoding)</li> <li>• AIFF, Audio Interchange File Format</li> </ul>	
<b>Digital Still Images</b>	<ul style="list-style-type: none"> <li>• TIFF (Tagged Image File Format)</li> <li>• GIF 89a Format</li> <li>• BMP Format</li> </ul>	<ul style="list-style-type: none"> <li>• JPEG 2000 Part 1 Format (JP2)</li> <li>• DNG Format</li> <li>• JFIF/JPEG Format</li> </ul>	<ul style="list-style-type: none"> <li>• TARGA Format</li> </ul>

# Metadata for Transfer

- The types of electronic records in use have changed and so have our needs.
  - There are more metadata rich formats.
  - We currently ask for indexes and system documentation.
  - More and more this exists as metadata that we can use to help automate processes.

# What are you using?

- We would like to work with you to identify:
  - Existing standards that you are using.
  - Your capabilities to export metadata that you use to search and retrieve e-records.
  - A “simplistic” set of metadata appropriate for transfer.

# Kevin De Vorse

Electronic Records Format Specialist

Policy Analysis and Enforcement

Office of the Chief Records Officer

Agency Services

National Archives and Records Administration



[kevin.devorse@nara.gov](mailto:kevin.devorse@nara.gov)

212-401-1631