METADATA
FOR PRESERVATION AND ACCESS

SMPTÉ BITS BY THE BAY | 2014
Seth Anderson, AVPreserve
METADATA
WHAT IS IT?
METADATA

Facts and statistics collected together for reference or analysis.

Referring to itself or to the conventions of its genre; self-referential.
data ABOUT data
“Metadata is structured information associated with an object for purposes of discovery, description, use, management, and preservation.”

- NISO, A Framework of Guidance for Building Good Digital Collections
Using Metadata to Catch a Whistleblower
Huffington Post (blog) - 1 hour ago

Metadata in 2013 was not a term widely-known to the general public. A quick definition might be that metadata is information about data—when and where the ...

Obama: Government Shouldn’t Hold Metadata in Bulk
NBCNews.com - 5 hours ago

The New Phone Metadata Program
In-Depth - Washington Post - Mar 25, 2014

White House unveils plan to end the NSA’s bulk collection of metadata
Blog - RT (blog) - 3 hours ago

Why Even Congress Supports Obama’s Proposal to End the NSA’s ...
In-Depth - Slate Magazine - Mar 25, 2014

NSA spying: phone "metadata" yields private details
Blog - SFGate (blog) - Mar 25, 2014

Explode in depth (1,427 more articles)

Republicans Prepare Bill to End NSA Phone Metadata Program
Mother Jones - 19 hours ago

Earlier this week President Obama announced a proposal to shut down the NSA’s bulk collection of telephone metadata. I thought that was great, but wondered ...

White House Fact Sheet on Metadata Proposal
Lawfare (blog) - 5 hours ago Share

White House Fact Sheet on Metadata Proposal … laid out in his
“Metadata is created throughout the lifecycle of an information object, includes more than just description”

- Andrea Leigh, Metadata Librarian, UCLA Film and Television Archive
Digital Content Life Cycle

- Describing
- Creating
- Managing
- Selecting
- Preserving
- Using & Reusing
- Discovering

http://beta.digipedia.org.uk/wiki/Digital_content_life_cycle
TYPES OF METADATA
Types

Reflect different uses and information needs

Where is the content?  Discovery Metadata
Who made it?    Discovery Metadata
Do I have permission to use it?  Rights Metadata
How can I display it properly?  Technical & Structural Metadata
How can I preserve it?  Preservation Metadata
Where did it come from?  Source Metadata
How does it relate to other digital objects?  Structural Metadata
Types

- Descriptive / Discovery
- Technical
- Rights
- Preservation
- Source
- Structural
Descriptive Metadata describes a resource for purpose of **identification** and **discovery**.
Descriptive Metadata describes a resource for purpose of **identification** and **discovery**.

- **World At Your Call, The**
  - Downloads: 1,766
  - Average rating: 3 stars (3 reviews)

- **Chevrolet Leader News (Vol. 1, No. 3)** - Handy (Jam) Organization
  - Promotional newsreels featuring a variety of Chevrolet-centric stories, including: heavy-duty trucks used in construction; St. Augustine, Florida, site of the oldest houses in America; old horseless carriage advertisements; a homemade streamlined house trailer in Guadalupe, California; auto interiors designed by New York jewelers; an Army convoy at Fort Sill, Oklahoma; and a Chihuahua dog standing in for a radiator ornament on a moving car in Grand Rapids, Michigan.
  - Keywords: Car culture; Newsreels; Automobiles; Advertising; Animals; Dogs
  - Downloads: 5,152
  - Average rating: 3 stars (2 reviews)

- **Coach for Cinderella, A** - Handy (Jam) Organization
  - Imaginative cartoon that breaks down the parts and features of a Chevrolet automobile into units that can be understood in terms of plant and animal life: caterpillars roll into circles and become automobile tires; in the "visible V-6" fireflies equal spark plugs and mice provide motive power. Naturally, tells part of the Cinderella story, too. Animation supervisor thought to be Frank Goldman.
  - Keywords: Animation; Advertising; Automobiles; Advertising; Folk tales
  - Downloads: 21,104
  - Average rating: 4 stars (12 reviews)

- **Girl on the Magazine Cover, The** - Handy (Jam) Organization
  - How professional models (and new automobiles) are photographed.
  - Keywords: Gender roles; Automobiles; Advertising; Media; Photography
  - Downloads: 15,160
  - Average rating: 3 stars (6 reviews)

- **Our Home Town: Doylestown, PA.** - Shad E. Graham
  - Advertising film presenting landmarks, businesses and people of a Philadelphia suburb.
  - Keywords: Philadelphia, Pa; Cities; U.S; Suburbia
  - Downloads: 5,375
  - Average rating: 3 stars (5 reviews)
**Technical Metadata**

Information related to **creation, display, playback, etc.**

### General

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete name</td>
<td>/Applications/Dropbox/Ukraine/Individual MP3s/Batuta.mp3</td>
</tr>
<tr>
<td>Format</td>
<td>MPEG Audio</td>
</tr>
<tr>
<td>File size</td>
<td>2.42 MiB</td>
</tr>
<tr>
<td>Duration</td>
<td>1mn 41s</td>
</tr>
<tr>
<td>Overall bit rate</td>
<td>192 Kbps</td>
</tr>
<tr>
<td>Album</td>
<td>Ethnomecanic</td>
</tr>
<tr>
<td>Track name</td>
<td>Batuta</td>
</tr>
<tr>
<td>Performer</td>
<td>Zdob Si Zdub</td>
</tr>
<tr>
<td>Recorded date</td>
<td>2006</td>
</tr>
<tr>
<td>Cover</td>
<td>Yes</td>
</tr>
<tr>
<td>Cover MIME</td>
<td>image/png</td>
</tr>
<tr>
<td>Comment</td>
<td><a href="http://www.mp3sugar.com/">http://www.mp3sugar.com/</a></td>
</tr>
</tbody>
</table>

### Audio

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>MPEG Audio</td>
</tr>
<tr>
<td>Format version</td>
<td>Version 1</td>
</tr>
<tr>
<td>Format profile</td>
<td>Layer 3</td>
</tr>
<tr>
<td>Mode</td>
<td>Joint stereo</td>
</tr>
<tr>
<td>Mode extension</td>
<td>MS Stereo</td>
</tr>
<tr>
<td>Duration</td>
<td>1mn 41s</td>
</tr>
<tr>
<td>Bit rate mode</td>
<td>Constant</td>
</tr>
<tr>
<td>Bit rate</td>
<td>192 Kbps</td>
</tr>
<tr>
<td>Channel(s)</td>
<td>2 channels</td>
</tr>
<tr>
<td>Sampling rate</td>
<td>44.1 KHz</td>
</tr>
<tr>
<td>Compression mode</td>
<td>Lossy</td>
</tr>
<tr>
<td>Stream size</td>
<td>2.33 MiB (96%)</td>
</tr>
</tbody>
</table>
## Rights Metadata

The Rights Metadata captures copyright information, permissions, licensing, etc.

<table>
<thead>
<tr>
<th>Time Created</th>
<th>2011:05:03 13:00:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Province City</td>
<td>DC</td>
</tr>
<tr>
<td>Country-Primary Location Name</td>
<td>USA</td>
</tr>
<tr>
<td>By-line</td>
<td>Lawrence Jackson</td>
</tr>
<tr>
<td>By-line Title</td>
<td>Official Photographer</td>
</tr>
<tr>
<td>Caption-Abstract</td>
<td>First Lady Michelle Obama greets students following a Flash Mob Dance at Alice Deal Middle School in Washington, D.C., May 3, 2011. (Official White House Photo by Lawrence Jackson). This official White House photograph is being made available only for publication by news organizations and/or for personal use printing by the subject(s) of the photograph. The photograph may not be manipulated in any way and may not be used in commercial or political materials, advertisements, emails, products, promotions that in any way suggests approval or endorsement of the President, the First Family, or the White House.</td>
</tr>
<tr>
<td>Copyright Notice</td>
<td>This official White House photograph is being made available only for publication by news organizations and/or for personal use printing by the subject(s) of the photograph. The photograph may not be manipulated in any way and may not be used in commercial or political materials, advertisements, emails, products, promotions that in any way suggests approval or endorsement of the President, the First Family, or the White House.</td>
</tr>
<tr>
<td>IPTC Digest</td>
<td>9e99fb9735a55610248863f7c0eb86c64</td>
</tr>
<tr>
<td>X Resolution</td>
<td>240</td>
</tr>
<tr>
<td>Y Resolution</td>
<td>inches</td>
</tr>
<tr>
<td>Displayed Units Y</td>
<td>inches</td>
</tr>
<tr>
<td>Global Angle</td>
<td>30</td>
</tr>
<tr>
<td>Global Altitude</td>
<td>30</td>
</tr>
<tr>
<td>Copyright Flag</td>
<td>True</td>
</tr>
<tr>
<td>Photoshop Quality</td>
<td>Standard</td>
</tr>
<tr>
<td>Photoshop Format</td>
<td>Standard</td>
</tr>
<tr>
<td>Progressive Scans</td>
<td>3 Scans</td>
</tr>
<tr>
<td>XMP Toolkit</td>
<td>XMP Core 4.1.1</td>
</tr>
<tr>
<td>Legacy IPTC Digest</td>
<td>9e99fb9735a55610248863f7c0eb86c64</td>
</tr>
<tr>
<td>Color Space</td>
<td>RGB</td>
</tr>
<tr>
<td>ID-Profile Name</td>
<td>sRGB IEC61966-2.1</td>
</tr>
<tr>
<td>City</td>
<td>Washington</td>
</tr>
<tr>
<td>State</td>
<td>DC</td>
</tr>
<tr>
<td>Country</td>
<td>USA</td>
</tr>
<tr>
<td>Category</td>
<td>P</td>
</tr>
<tr>
<td>Authors Position</td>
<td>Official Photographer</td>
</tr>
<tr>
<td>Credit</td>
<td>The White House</td>
</tr>
<tr>
<td>Source</td>
<td>The White House</td>
</tr>
<tr>
<td>Supplemental Categories</td>
<td>Public Schedule</td>
</tr>
<tr>
<td>Date Created</td>
<td>2011:05:03 13:00:00</td>
</tr>
<tr>
<td>Modify Date</td>
<td>2011:05:03 13:00:00</td>
</tr>
<tr>
<td>Create Date</td>
<td>2011:05:03 13:00:00</td>
</tr>
</tbody>
</table>
Rights Metadata captures copyright information, permissions, licensing, etc.
Preservation Metadata

Information about the creation of the asset, integrity management, dependencies, preservation actions, etc.

- Source
- Assessment/Inspection/Preparation
- Treatment
- Transfer
- Pres Master 1
- Pres Master 2
- Access Master 1
- Access Master 2
- Transcode
- Access Copy 1
- Access Copy 2
- Access Copy 3

who?
when?
what equipment?
outcome?
Preservation Metadata

Information about the creation of the asset, integrity management, dependencies, preservation actions, etc.

who?
when?
integrity check?
Preservation Metadata
Information about the creation of the asset, integrity management, dependencies, preservation actions, etc.

MD5 = d214ea9efb6ef1d8d3b3aaf0b612f151
Source / Physical Metadata
describes the provenance of an asset digitized from an analog source

Welles_war_worlds.mp3
Structural Metadata

Information about how compound objects are put together, describes the relationship between a digital object and component parts.
• Enable discovery, search, browse, retrieval

• Collection management

• Context for users

• Preservation planning

• Obsolescence monitoring

• Rights management

• Provenance / Event history

• Collection profiling

• Authenticity and integrity management
META DATA
STANDARDS
Types of Standards

Content Standards (rules)
Value Standards (authorities)
Structure Standards (schemas)

good metadata practice uses all three
Content Standards

Rules or best practice guidelines for how descriptive metadata is expressed.

TITLE

Provide either a formal title or a concise supplied title in accordance with the rules of multilevel description and national conventions.

If appropriate, abridge a long formal title, but only if this can be done without loss of essential information.

For supplied titles, at the higher level, include the name of the creator of the records. At lower levels one may include, for example, the name of the author of the document and a term indicating the form of the material comprising the unit of description and, where appropriate, a phrase reflecting function, activity, subject, location, or theme.

Distinguish between formal and supplied titles according to national or language conventions.

ISAD(G):
Content Standards

Rules or best practice guidelines for how descriptive metadata is expressed.

- AACR2
- RDA
- DACS
- ISAD(G)
- CCO
- FIAF Cataloging Rules
Value Standards (Authorities & Thesauri)

Controlled vocabularies, lists, or thesauri used to populate data structure standards or element sets

World War I

World War, 1914-1918

Great War

WWI

Library of Congress Subject Headings

First World War
Structure Standards
Schemas and/or element / field sets and rules for their use

Structures / Schemas define:

- **Data labels**: Element or field names
- **Definition** of the field or element
- **Constraints & Rules** (repeatable? required? optional? character length?)
Structure Standards
Schemas and/or element / field sets and rules for their use

Descriptive MD Standards
MARC21
Dublin Core
MODS
PBCore / EBU Core
AES-60 Core audio metadata
ID3
Schema.org

User Communities / Resource Type
General - Libraries
General
Moving Image & Audio
Audio / Broadcasting
mp3
Web
Structure Standards
Schemas and/or element / field sets and rules for their use

Technical MD Standards
MIX
PBCore
AES
audioMD

User Communities / Resource Type
Still Images
Moving Image and Audio
Audio / Broadcasting
Audio
“Some file formats enable the capture of technical, and other, metadata within their files, which has the advantage of keeping the files self-descriptive. However, by extracting and storing metadata explicitly we may also benefit.”

- ISQ Special Issue: Digital Preservation, Spring 2010, p.11
Structure Standards
Schemas and/or element / field sets and rules for their use

Preservation MD Standards
PREMIS  Preservation Metadata Implementation Strategies
AES-57  Audio object structures for preservation and restoration

User Communities / Resource Type
General  Audio
Structure Standards
Schemas and/or element / field sets and rules for their use

Rights MD Standards

METSRights Rights Declaration MD
ODRL Open Digital Rights Language
Creative Commons

User Communities / Resource Type
General
General
General
Structure Standards
Schemas and/or element / field sets and rules for their use

Metadata Wrapper / Container Standards
METS Metadata Encoding and Transmission Standard

User Communities / Resource Type
General

<!--XML Generated at Fri Nov 11 20:56:48 -0500 2011 by rails pbcore database-->
<!--created at 2008-10-22 13:39:10 UTC by unknown-->
<!--updated at 2008-10-22 13:39:10 UTC by unknown-->

<pbcoreIdentifier source="PBS PODS">FHUS_000102</pbcoreIdentifier>
<pbcoreIdentifier source="pbcore XML database UUID">e6d58245-41e2-4834-b375-d29a375e5c7e</pbcoreIdentifier>
<pbcoreTitle titleType="Series">Freedom: A History of US</pbcoreTitle>
<pbcoreTitle titleType="Episode">Revolution</pbcoreTitle>
<pbcoreSubject source="User Keywords">Don Dohler; sci fi; science fiction</pbcoreSubject>
<pbcoreDescription descriptionType="Segment">Revolution: Colonial Americans fought together to defeat the world's most awesome military power. Now they strive to create a new kind of government that will live up to their high ideals. George Washington and Thomas Jefferson become presidents; Lewis and Clark undertake their great expedition; and the American Dream is born.</pbcoreDescription>
<pbcoreGenre source="PBS PODS">History</pbcoreGenre>
<pbcoreRelation>
  <pbcoreRelationType>Is Part Of</pbcoreRelationType>
  <pbcoreRelationIdentifier>FHUS</pbcoreRelationIdentifier>
</pbcoreRelation>
<pbcoreCoverage coverage="USA">USA</pbcoreCoverage>
<pbcoreCoverage coverageType="Spatial">Spatial</pbcoreCoverage>
<pbcoreAudienceRating audienceRating="General Audience">General Audience</pbcoreAudienceRating>
<pbcoreCreator>
  <creator>KUNHARDT PRODUCTIONS, INC.</creator>
  <creatorRole>Production Unit</creatorRole>
</pbcoreCreator>
<pbcoreContributor>
  <contributor>Co, Katie</contributor>
  <contributorRole>Host</contributorRole>
</pbcoreContributor>
<pbcoreContributor>
  <contributor>Foner, Eric</contributor>
  <contributorRole>Other</contributorRole>
</pbcoreContributor>
<pbcorePublisher>
  <publisher>Thirteen/WNET</publisher>
  <publisherRole>Presenter</publisherRole>
</pbcorePublisher>
<pbcorePublisher>
  <publisher>Thirteen/WNET</publisher>
  <publisherRole>Copyright Holder</publisherRole>
</pbcorePublisher>
<pbcorePublisher>
  <publisher>PBS-PLUS</publisher>
  <publisherRole>Distributor</publisherRole>
</pbcorePublisher>
<pbcoreRightsSummary>
  Standard Group; 4 Releases, 1/3/2003 to 1/3/2006; School: Based on Each Broadcast, 12 months
</pbcoreRightsSummary>
</pbcoreDescriptionDocument>
Embedded metadata

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 1.0. Approved version, September 15, 2009 (clarification wording November 24, 2009)

TABLE OF CONTENTS

I. THE BEXT CHUNK
   I.A. Strongly recommended elements (page 2)
        Originator, OriginatorReference, Description (for additional identifiers),
        OriginationDate, Version, Reserved
   I.B. Recommended element (page 5)
        TimeReference
   I.C. Optional elements (page 5)
        OriginationTime, CodingHistory

II. THE INFO CHUNK
    II.A. Recommended element (page 7)
         IARL (archival location)
    II.B. Optional elements (page 7)
         INAM (name, title), ICMT (comment for identifiers), ICRD (creation date), ICOP
         (copyright and other restrictions)

III. SPECIFICATIONS AND OTHER REFERENCES (page 9)

What is this document?
This is one of three documents pertaining to the embedding of metadata in digital audio files prepared by the Federal Agencies Audio-Visual Working Group in 2009. The two companion documents are:

- Introductory Discussion for the Proposed Federal Agencies Guideline
- Consultant’s report on embedding options in digital audio files

BWF MetaEdit

http://sourceforge.net/projects/bwfmetaedit/

<table>
<thead>
<tr>
<th>Coding algorithm</th>
<th>Sampling frequency</th>
<th>Bit rate</th>
<th>Word length</th>
<th>Mode</th>
<th>Text, free string</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCM</td>
<td>96000</td>
<td>24</td>
<td>stereo</td>
<td></td>
<td>original</td>
</tr>
<tr>
<td>PCM</td>
<td>44100</td>
<td>16</td>
<td>stereo</td>
<td></td>
<td>Wavelab 6.10; UV22HR dither</td>
</tr>
</tbody>
</table>

Save   Tech  Core
- OrignationTime: 15:25:54
- TimeReference (translated): 00:00:10.628
- BextVersion: 468738
- UMID: 0

A=PCM,F=96000,W=24,M=stereo,T=original
A=PCM,F=44100,W=16,M=stereo,T=original

Vito Acconci
8/1/08
General
Complete name : /Users/kvanmalssen/Documents/AVPS/metro/filemgmt/oct2010/handout/Description/Full/full1.wav
Format : Wave
File size : 5.34 MiB
Duration : 21s 88ms
Overall bit rate : 2123 Kbps
Track name : 3 Columns for America
Director : Vito Acconci
Producer : Museum of Modern Art
Description : Vito Acconci / 1976 / 3 Columns for America / Installation work includes table, 3 stools and a blackboard
Recorded date : 8/1/08 15:25
Encoded date : UTC 2008-01-08 15:25:54
Writing application : WaveLab 6.10
Encoding settings : A=PCM,F=96000,W=24,M=stereo,T=original / A=PCM,F=44100,W=16,M=stereo,T=Wavelab 6.10; UV22HR dither
Original source form : Audiocassette

Audio
ID : 0
Format : PCM
Codec ID : 1
Codec ID/Hint : Microsoft
Duration : 21s 88ms
Bit rate : 2117 Kbps
Channel(s) : 2 channels
Sampling rate : 44.1 KHz
Resolution : 24 bits
Stream size : 5.32 MiB (100%)
### MDQC Tool

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Tag</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Originator</td>
<td>[ignore Tag]</td>
<td>Bext Originator</td>
</tr>
<tr>
<td>Originator Reference</td>
<td>[ignore Tag]</td>
<td>Bext OriginatorReference</td>
</tr>
<tr>
<td>Product</td>
<td>[ignore Tag]</td>
<td>LIST–INFO IPRD (Product (Album))</td>
</tr>
<tr>
<td>Relative Timestamp</td>
<td>[ignore Tag]</td>
<td>1:04:04.444</td>
</tr>
<tr>
<td>Release Date</td>
<td>[ignore Tag]</td>
<td>XMP Release Date</td>
</tr>
<tr>
<td>Rights</td>
<td>[ignore Tag]</td>
<td>XMP Copyright Notice</td>
</tr>
<tr>
<td>Sample Rate</td>
<td>[ignore Tag]</td>
<td>96000</td>
</tr>
<tr>
<td>Scene</td>
<td>[ignore Tag]</td>
<td>XMP Video Data Scene</td>
</tr>
<tr>
<td>Shot Date</td>
<td>[ignore Tag]</td>
<td>XMP Video Data Date Shot</td>
</tr>
<tr>
<td>Shot Name</td>
<td>[ignore Tag]</td>
<td>XMP Video Data shot/take</td>
</tr>
<tr>
<td>Software</td>
<td>[ignore Tag]</td>
<td>LIST–INFO ISFT (Software)</td>
</tr>
<tr>
<td>Source</td>
<td>[ignore Tag]</td>
<td>XMP Origin Source</td>
</tr>
<tr>
<td>Source Form</td>
<td>[ignore Tag]</td>
<td>LIST–INFO ISRF (Source Form)</td>
</tr>
<tr>
<td>Startup Profile</td>
<td>[ignore Tag]</td>
<td>Print</td>
</tr>
<tr>
<td>State</td>
<td>[ignore Tag]</td>
<td>XMP State</td>
</tr>
</tbody>
</table>
FUTURE

dirty data

VOCAB

clean data

MODEL

RULES