open.michigan

Author(s): Paul Conway, 2008-2011.

License: Unless otherwise noted, this material is made available under the terms of the **Creative Commons Creative Commons Attribution - Non-Commercial - Share Alike 3.0 License**: http://creativecommons.org/licenses/by-nc-sa/3.0/

We have reviewed this material in accordance with U.S. Copyright Law **and have tried to maximize your ability to use, share, and adapt it.** The citation key on the following slide provides information about how you may share and adapt this material.

Copyright holders of content included in this material should contact **open.michigan@umich.edu** with any questions, corrections, or clarification regarding the use of content.

For more information about **how to cite** these materials visit http://open.umich.edu/education/about/terms-of-use.

Any **medical information** in this material is intended to inform and educate and is **not a tool for self-diagnosis** or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. Please speak to your physician if you have questions about your medical condition.

Viewer discretion is advised: Some medical content is graphic and may not be suitable for all viewers.

University of Michigan

CC BY NC SA

Citation Key

for more information see: http://open.umich.edu/wiki/CitationPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV **Public Domain – Government**: Works that are produced by the U.S. Government. (17 USC § 105)

PD-EXP **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

CC ZERO **Creative Commons – Zero Waiver**

CC BY **Creative Commons – Attribution License**

CC BY-SA **Creative Commons – Attribution Share Alike License**

CC BY-NC **Creative Commons – Attribution Noncommercial License**

CC BY-NC-SA **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

© FAIR USE **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

SI 675 Digitization for Preservation

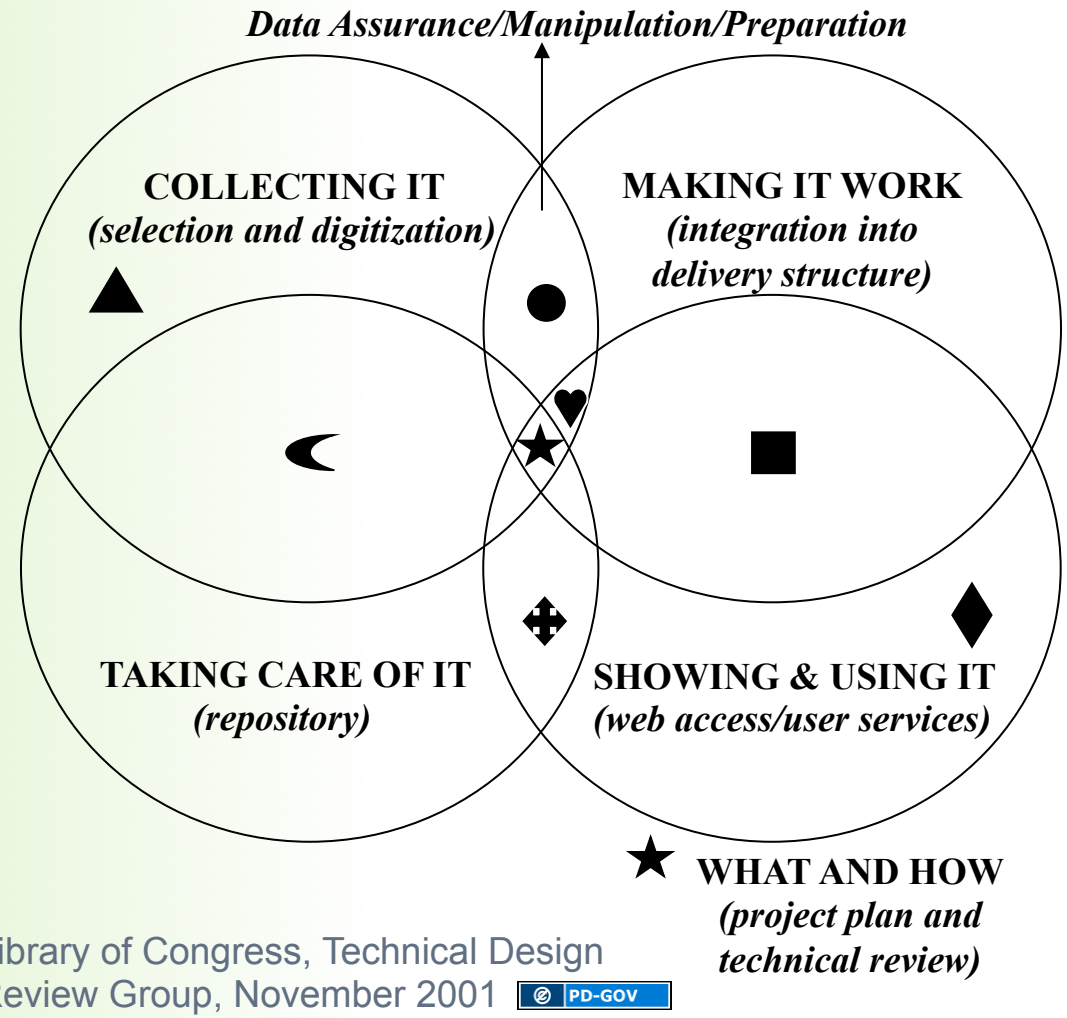
Week 7 – Metadata for Image Objects

Outline

- Managing a digitization program - debrief
- Metadata for images
- File formats

Yad Vashem and Google Partner to Preserve and Share Holocaust Archives:
http://www.yadvashem.org/
Search on Yecheskel Fleischer

Aspects of Digital Collection Creation and Maintenance



ROLES

★	*Production Coordinator* *Technical Review Group*
▲	*Content Custodian* *Conservator* *Capture Specialist* *Copyright Researcher*
♥	*Data Wrangler*
●	*Description Creator* *Quality Assurance Specialist* *Editor*
■	*Applications Developer*
♦	*Graphic Interface Designer*
✥	*Systems Engineer*
☾	*Digital Custodian*

Library of Congress, Technical Design Review Group, November 2001 PD-GOV

Applying Standards in Practice

- Analogy: pieces of a complex puzzle
  - Edge pieces provide a framework
  - Connections among similar functions and concepts
  - Still some missing pieces, but not so many that the overall picture can't be discerned
- Standards issues range from well-defined to unknown
  - Product of digitization increasingly standardized
  - Matching standards to workflow fairly well understood
  - Impact of decision making marginally clear
  - User requirements not well understood
- Preservation: from replacement to transformative use

Metadata Functions in Digitization

- Describe objects
  - Original, surrogate
- Structure relationships
  - Internal sequencing
  - External context
- Manage life cycle
  - Origins, rights
  - Technical characteristics
  - Preservation (changes)
  - Location

Metadata Standards

- Making of America II
  - Descriptive [about object & source]
  - Structural [internal & external]
  - Administrative [technical + preservation]
- Library of Congress – Standards Development Office
  - http://www.loc.gov/standards/
- Metadata for digital content (2009)
  - Descriptive elements for bitmaps
  - http://www.loc.gov/standards/mdc/elements/

Metadata for Image Collections

- Dublin Core is minimum for description
  - http://dublincore.org/
- Technical and administrative metadata are in a state of flux
  - MIX
  - PREMIS and METS record
  - Specialized, local metadata schemas

Technical Metadata for Images

- Origins:
  - *Automatic Exposure: RLG-led initiative to promote technical metadata*
    - *http://www.oclc.org/research/activities/past/rlg/automaticexposure/default.htm*
  - *NISO Z39.87: Data Dictionary—Technical Metadata for Digital Still Images*
    http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c
  - See handout of metadata elements

- MIX: Metadata for Images in XML…
  - http://www.loc.gov/standards/mix/
- Uses: Harvard JHOVE
  - Detects formats and assesses how well they conform to standards
  - JHOVE - JSTOR/Harvard Object Validation Environment

ANSI/NISO Z39.87-2006 – Object Identifier

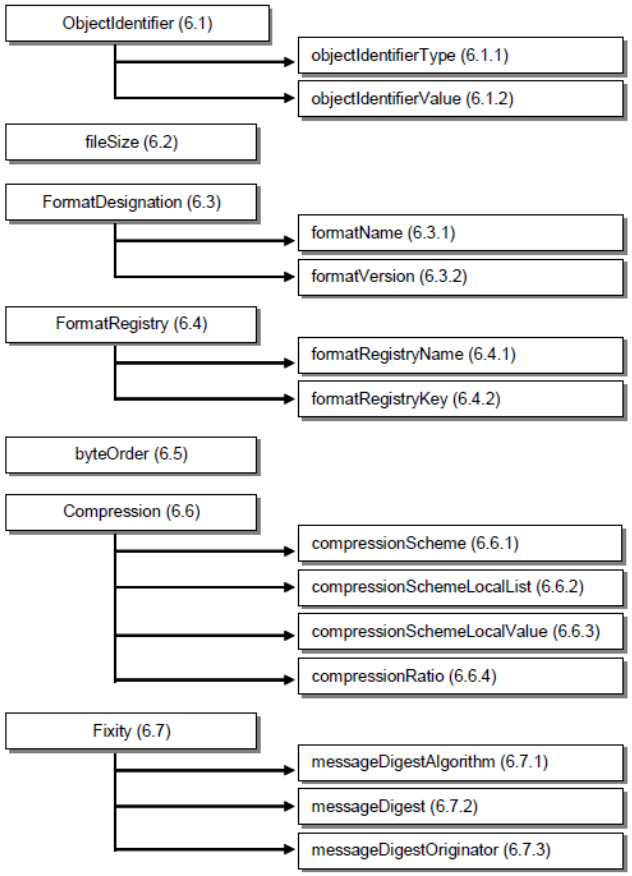


Figure 1: Logical structure of basic digital object information

6.3.1 formatName

Definition	*a data element that designates the format name or description of the file format*
Type	string
Obligation	M
Repeatable	N
Values (Examples)	image/jp2 image/geotiff Adobe PDF base64
Notes	This data element is drawn from the PREMIS data element set. Values should be taken from a controlled vocabulary. It is permissible to either list proper format names (e.g., "Adobe PDF") or MIME types (e.g., "image/tiff" or "image/jp2")
Use	Manager System User

PD-INEL 2006 by the National Information Standards Organization.

ANSI/NISO Z39.87-2006 – Basic Characteristics

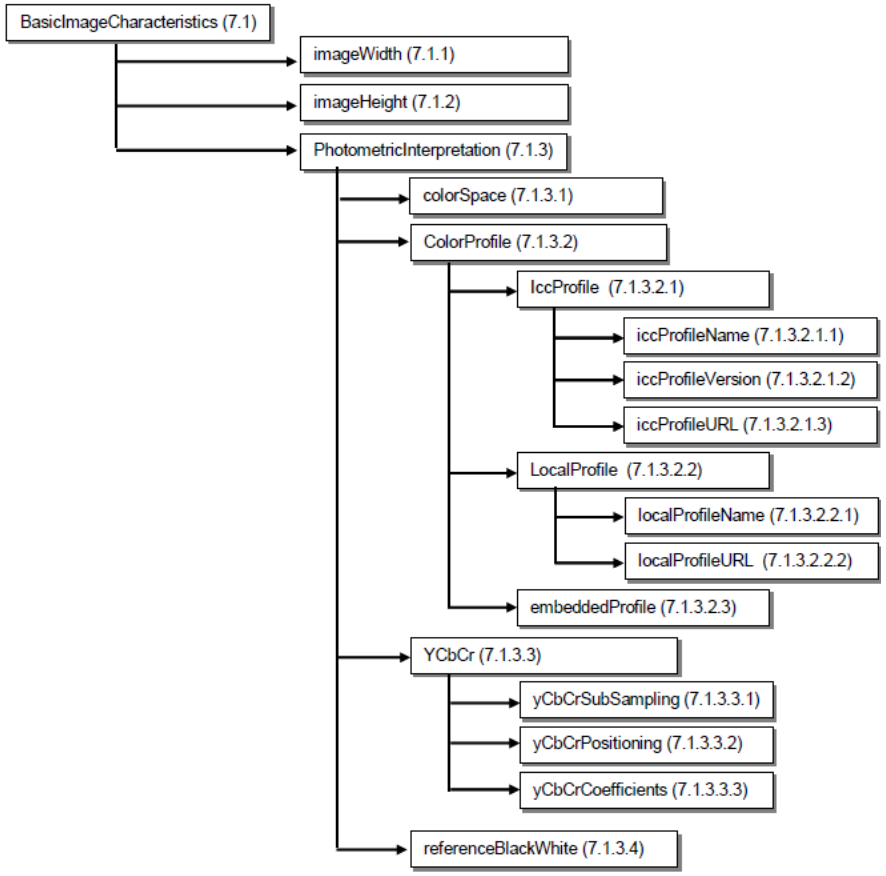


Figure 2: Logical structure of BasicImageCharacteristics

7.1.2 imageHeight

Definition	*a data element that specifies the height of the digital image, i.e. vertical or Y dimension, in pixels*
Type	positive integer
Obligation	M
Repeatable	N
Values	2130
(Examples)	1200
Notes	The image height may be the shorter or longer dimension of the image, depending upon the orientation of the camera or scanner during image capture. For multiple-resolution image file formats, value shall specify the highest resolution. This field may be used to calculate 9.1.2.2 ***ySamplingFrequency*** when 8.1.3.2.2 ***sourceYDimensionUnit*** is in inches and 9.1.2 ***samplingFrequencyUnit*** = 2. Formula to calculate ***ySamplingFrequency***: ***ySamplingFrequency = imageHeight/sourceYDimension***
Use	System (required field for image viewers [size]) Manager (one of the quantitative metrics to evaluate image quality) User

PD-INEL 2006 by the National Information Standards Organization.

ANSI/NISO Z39.87-2006 – Source Info

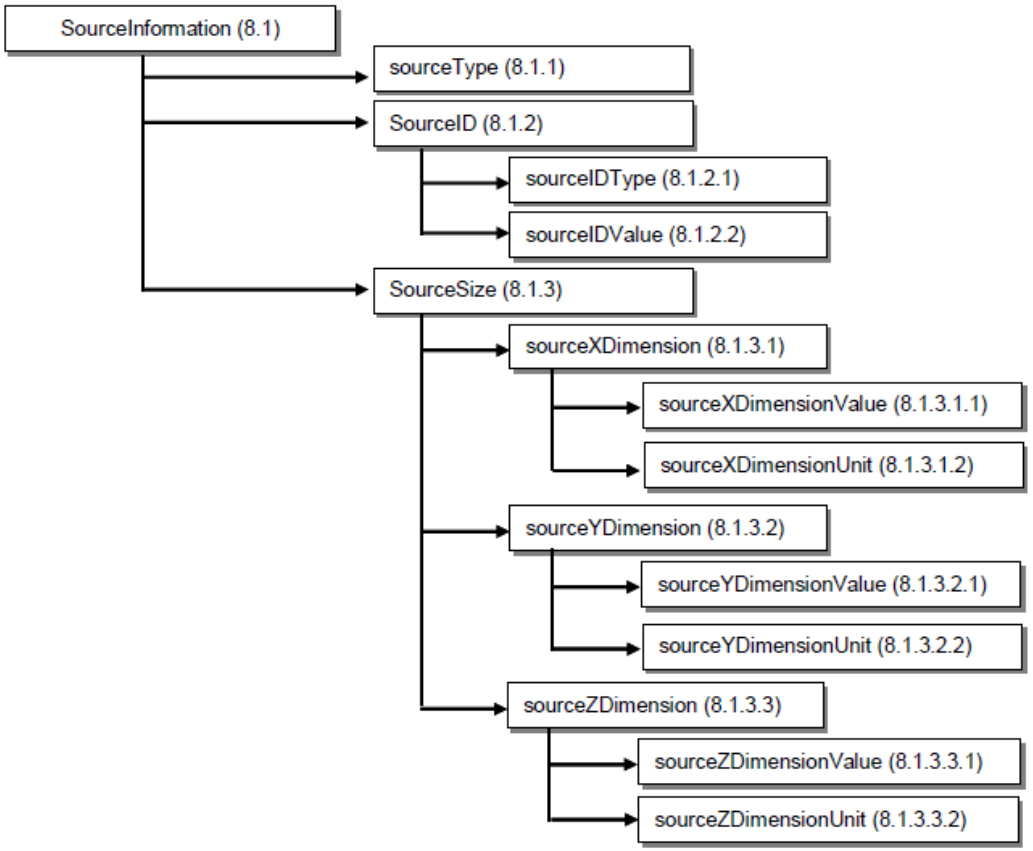


Figure 4: Logical structure of SourceInformation

cifies the width of the imaged object

f the source and the dimensions actually imaged
nsion of area imaged.

graphic print measuring 8 inches in the X dimension
the image file's ***sourceXDimensionValue*** would

c print is imaged in its entirety with a 1-inch border
ile's ***sourceXDimensionValue*** would be "10",
n of the total area imaged includes the space

ge map is imaged, and the detail measures
sion while the map measures 40 inches in the
file's ***sourceXDimensionValue*** would be "6".

v is imaged, and the window measures 6 feet in
he church wall in which it is located has a
t in the X dimension, the image file's
ue would be "6".

hes, meters, etc.) used must be specified in
nsionUnit.

l to record, the ***sourceXDimensionValue*** may be
SamplingFrequencyPlane.

PD-INEL 2006 by the National Information Standards Organization.

MIX: Metadata for Images in XML

- MIX Schema Version 2.0 (current version)
  - Implements ANSI/NISO Z39.87 – 2006
  - Standard maintained by Library of Congress
  - http://www.loc.gov/standards/mix/

```
<xsd:schema targetNamespace="http://www.loc.gov/mix/v20"
            xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            xmlns="http://www.loc.gov/mix/v20"
            elementFormDefault="qualified"
            attributeFormDefault="unqualified"
            >

  <xsd:element name="mix">
    <xsd:annotation>
      <xsd:documentation>NISO Data Dictionary Technical Metadata for Digital Still Images (May 1, 2006)</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:complexContent>
        <xsd:extension base="mixType"/>
      </xsd:complexContent>
    </xsd:complexType>
  </xsd:element>
```

MIX Code for Z39.87 – 7.1.2 Image Height

ANSI/NISO “Container” = MIX “complexType” with “elements”

```
<!-- BasicImageInformationType -->
<xsd:complexType name="BasicImageInformationType">
  <xsd:sequence>
    <xsd:element name="BasicImageCharacteristics" minOccur
      <xsd:annotation>
        <xsd:documentation>NDD id="7.1"</xsd:documentation
      </xsd:annotation>
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="imageWidth" type="positiveInt
            <xsd:annotation>
              <xsd:documentation>NDD id="7.1.1"</xsd:documentation>
            </xsd:annotation>
          </xsd:element>
          <xsd:element name="imageHeight" type="positiveIntegerType" minOccurs="0" maxOccurs="1">
            <xsd:annotation>
              <xsd:documentation>NDD id="7.1.2"</xsd:documentation>
            </xsd:annotation>
          </xsd:element>
```

```
<!-- NISO Section 7 -->
-<BasicImageInformation>
 -<BasicImageCharacteristics>
    <imageWidth>400</imageWidth>
    <imageHeight>200</imageHeight>
   -<PhotometricInterpretation>
      <colorSpace/>
     -<ColorProfile>
       -<IccProfile>
```

MIX 2.0: http://www.loc.gov/standards/mix/mix20/mix20.xsd

i3a: International Imaging Industry Association

- IT 10: Electronic Still Picture Imaging
  - Picture Transfer Protocol
  - ISO 15740:2005

- International standard for exchange of images and metadata from 95% of cameras produced in the world.
  - One standard for USB
  - One standard for TCP/IP

- Platform independent
  - Windows Media Transport Protocol; Mac OS X; Linux

MIX Uses

- Adobe Extensible Metadata Platform (XMP)
  - Modifies scanner control software for metadata capture
  - Example: PhotoShop "File Info…"
  - http://www.adobe.com/products/xmp/overview.html

- Harvard JHOVE
  - Detects formats and assesses how well they conform to standards
  - JHOVE - JSTOR/Harvard Object Validation Environment

File Formats

- TIFF – Tagged Image File Format
  - http://www.awaresystems.be/imaging/tiff/tifftags/baseline.html
- PNG – Portable Network Graphics
  - ISO/IEC 15948
  - http://www.libpng.org/pub/png/
- JPEG 2000
  - http://www.jpeg.org/jpeg2000/index.html
- Benefits of JPEG 2000
  - http://www.digitizationguidelines.gov/stillimages/presentations.html

- Potential use cases for JHOVE include:
- Identification
  - "I have an object; what format is it?"
- Validation
  - "I have an object that purports to of format *F*; is it?"
  - "I have an object of format *F*; does it meet profile *P* of *F*?"
  - "I have an object of format *F* and external metadata about *F* in schema *S*; are they consistent?"
- Characterization
  - "I have an object of format *F*; what are its salient properties (given in schema *S*)?"

JHOVE: http://hul.harvard.edu/jhove/

Summary of Key Concepts

- Digitization can be a preservation strategy, under certain circumstances
- Digitization is representation of an artifact in digital form
  - Digital coding
  - Extensive overt and subtle decision making in workflow
- Digitization for preservation depends on developments in image science and evolving best practices
- Targets provide confidence that scanning equipment is performing to expectations
- Use of technical metadata is essential to support preservation goals

Thank you!

Paul Conway
Associate Professor
School of Information
University of Michigan
www.si.umich.edu