

# Introduction to Digital Preservation for Moving Image and Sound

Instructor: Lauren Sorensen (she/her)

[laurens@nyu.edu](mailto:laurens@nyu.edu)

<http://laurensorensen.info>

Twitter: @laurensx

# Day 1

- Introductions: name, location, institution, why you're here
- Format for workshop:
  - Camera on is optional; unmuting yourself is fine, but please use "raise hand" so we know in advance.
  - In general, follow process of digitization following accessioning: moving image and sound specific first, then digital preservation.
  - Use the question box to ask questions along the way, and we will save time at the end for live questions
- Topics day 1: Vocab, identification, risk factors, digitization process, digital preservation activities and concepts
- Topics day 2: Metadata, software, storage, consultations
- Exercises day 1: Discussion, MD5 checksum creation
- Exercises day 2: Discussion, NDSA levels

# Quick vocabulary review

Codec

Wrapper aka Container

Analog

Digital

Emulation

Migration

Fixity

Ingest

Transcode

Lossless

Lossy

Generation

Digital repository

Open source

Proprietary

Checksum

Compression

Scripts

# Identification: Analog versus digital

## Analog examples:

### Video

- 1/2" open-reel
- 3/4" Umatic
- VHS, S-VHS

### Film (any kind)

- 16mm
- 8mm
- Super8mm
- 35mm
- 70mm

### Audio

- 1/4" open reel
- Audiocassette
- Microcassette

## Digital examples:

### Video

- MiniDV
- DV
- DVD / DVD-R

### Audio

- DAT
- MiniDisc
- CD / CD-R



# Film



35mm



16mm



8mm



Super 8

# Video

## ► 1/2" Open Reel



TOP ►



BOTTOM ►

---

---

**FORMAT NAME** ► 1/2" Open Reel

---

---

---

---

**ANALOG OR DIGITAL** ► Analog

---

---

---

---

**DATE INTRODUCED** ► 1965

---

---

---

---

**DATES IN USE** ► 1965 – late 1970s

---

---

---

---

**TAPE WIDTH** ► 1/2"

---

---

**REEL DIMENSIONS** ► Commonly used reels are 5" in diameter or 7 1/4" in diameter.

**TAPE CONTAINER** ► The most common containers for smaller tapes are approximately 5 1/2" x 5 1/2" x 1" and made of a soft or hard plastic. Larger reels will be in containers approximately 8 3/8" x 8 3/8" x 1 1/4". The plastic container may also be inside of a separate paperboard sleeve. There are other containers that are approximately the same size but have handles.

# Video

SAFE MODE ►



RECORD  
MODE ►



Record Mode

# Audio



Preservation Self-Assessment Program: Collection ID Guide [PSAP](#)) | [Collection ID Guide](#)

Image 1/4' audio: <https://www.flickr.com/photos/124076687@N04/14107107418/>

Audiocassette image: <https://pixabay.com/photos/cassette-tape-plastic-tape-audio-164396/>

# Digitization priorities, risk factors

- Basic inventory / accessioning at item-level
  - Fields: ID number, title, primary author/artist, format, notes on condition, “generation”
- Identification, evaluation & inspection
  - Preservation Self Assessment Program (PSAP) Collection ID guide: <https://psap.library.illinois.edu/collection-id-guide>
  - Texas Commission on the Arts: Videotape Identification and Assessment Guide: <https://www.arts.texas.gov/wp-content/uploads/2012/04/video.pdf>
- Analog video: prioritize based on age and condition
  - Equipment obsolescence. Sticky-shed syndrome, soft binder syndrome. Tapes from late 70s to early 80s esp at risk.
- Analog audio: ¼ - inch open-reel tape, micro-cassette, audiocassette.
  - Equipment obsolescence. Sticky-shed syndrome, soft binder syndrome. Tapes from late 70s to early 80s esp at risk.
- Tape-based digital media: for example, DAT, DV, MiniDV.
  - Equipment obsolescence esp with DAT, extremely fragile base on MiniDV tape, “metal evaporate” or ME.
- Film: Super8mm, 8mm, 16mm, 35mm, 9.5mm, 70mm.
  - Risks: vinegar syndrome for acetate film, flammability for nitrate, brittleness and breaking.
- Optical media: DVDs, DVD-Rs, CDs, CD-Rs, MiniDiscs.
  - Risks: Scratching, dye fading for recordable discs, ink seeping through, obsolescence.

Questions so far

15 min break

# Digitization techniques & example specifications

- Video & Audio:
  - Analog: Playback deck → capture card & cables → capture software → hard drive
  - Digital: firewire or USB → capture software → hard drive
- Film
  - Scanner → capture card → capture software → hard drive

Example target formats:

California Revealed: [California Revealed Statement of Work for Audiovisual Materials 2020-2021 Project Description California Revealed \(CA-R\)](#)

NYPL Preservation: <https://nypl.github.io/ami-preservation/pages/ami-specifications.html>



# Digitization techniques & example specifications cont'd

Bay Area Video Coalition specs:

## **Analog Video (NTSC)**

Video Codec: 10-Bit Uncompressed (we also do FFV1, and everything else is the same except the MKV wrapper and the data rate is variable)

Wrapper: MOV (or MKV for FFV1)

Frame Size: 720x486

Frame Rate: 29.97

Display Aspect Ratio: 4:3

Pixel Aspect Ratio: 0.9

Video Bit Depth: 10

Video Bit Rate: 224 Mbps

Color Matrix: BT.601

Color Space: YUV

Chroma Subsampling: 4:2:2

Audio Codec: Linear PCM

Audio Sampling Rate: 48kHz

Audio Bit Depth: 24

## **Digital Video (NTSC)**

Video Codec: Native DV wrapped in MOV

Wrapper: MOV

Frame Size: 720x489

Frame Rate: 29.97

Display Aspect Ratio: 4:3 (or 16:9 if anamorphic)

Pixel Aspect Ratio: 0.9 (ot 1.2 if anamorphic)

Video Bit Depth: 8

Video Bit Rate: 24.4 Mbps

Color Space: YUV

Chroma Subsampling: Whatever the tape was recorded as (typically 4:1:1)

Audio Codec: Linear PCM

Audio Sampling Rate: Whatever the tape was recorded as

Audio Bit Depth: Whatever the tape was recorded as

# Digitization techniques & example specifications cont'd

Bay Area Video Coalition specs:

## **Analog Audio**

Sampling Rate: 96kHz

Bit Depth: 24

Codec: Linear PCM

Wrapper: Broadcast WAVE

## **Digital Audio**

Sampling Rate: Whatever the tape was recorded as

Bit Depth: Whatever the tape was recorded as

Codec: Linear PCM

Wrapper: Broadcast WAVE

# Vendors: questions to ask

- Look for referrals from AV archivists [AMIA, ARSC listservs]
- What formats they handle, what experience they have
- Concerns about digitization, any flaking, smell or deterioration you've noticed
- Insurance to cover material while in their possession, during transport
- How originals will be stored while in their possession
- Description of inspection and cleaning procedure, is there a plan for tracking any damage upon receipt
- File formats they can produce fits your specifications, any workflow questions
- Technical or procedural metadata you would like them to provide
- Can they match the specifications you've laid out for a SIP?
- What quality control or assurance is offered
- If analog video, do they incorporate a time-base corrector?
- Delivery using Bag specification, or other way of documenting file checksums

## In-house: questions to ask

- What technical expertise do you have in-house and what can be outsourced?
- What equipment can you source from your organization, eBay or otherwise?
- Can you identify deteriorated materials that may need special treatment? E.g. mold or vinegar syndrome?
- What file formats software can produce fits your specifications?
- Technical or procedural metadata you would like to capture, and in what schema / format
- If analog video, do you have and know how to use a time-based corrector?
- Delivery of digitized content, including SIP package to repository, what happens to it and how is it managed going forward?

# Born digital: setting up a workstation

- Anti-static mat for table, your wrist; “clean” area in case hard drives need to be manipulated or taken out of their housing
- Write-blocker: Tableau brand
- Necessary cables and input / outputs given what connections original media has (USB, firewire, etc)
- Software:
  - consider Bitcurator, with caveat that using it as a VM may be tricky, since AV is resource intensive on CPUs.
  - BagIt / Bagger depending on operating system
  - Oxygen XML (\$) / BBEdit

# Resource page: target formats and born digital

**Film:** NARA, “Digital Moving Images from Film-based Source Material”

<https://www.archives.gov/preservation/products/reformatting/mopix-digital.html>

**Video:** <https://nypl.github.io/ami-preservation/pages/ami-specifications.html>

**Audio:** ARSC Guide to Audio Preservation: <https://www.clir.org/wp-content/uploads/sites/6/pub164.pdf>

- IASA TC-O4 document - outlines audio preservation format and specs in detail: <https://www.iasa-web.org/tc04/audio-preservation>

**Born Digital:** FADGI Born Digital Video working group,

[http://www.digitizationguidelines.gov/guidelines/video\\_bornDigital.html](http://www.digitizationguidelines.gov/guidelines/video_bornDigital.html)

Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house (Ricky Erway, OCLC): <https://www.oclc.org/content/dam/research/publications/library/2013/2013-02.pdf>

Library of Congress, “Preserving Write-Once DVDs Producing Disc Images, Extracting Content, and Addressing Flaws and Errors”: <http://www.digitizationguidelines.gov/audio-visual/documents/Preserve DVDs BloodReport 20140901.pdf>

Elvia Arroyo, “Tell Us about Your Digital Archives Workstation” A Survey and Case Study

[http://www.ala.org/alcts/sites/ala.org.alcts/files/content/\\_Tell%20Us%20about%20Your%20Digital%20Archives%20Workstation\\_.pdf](http://www.ala.org/alcts/sites/ala.org.alcts/files/content/_Tell%20Us%20about%20Your%20Digital%20Archives%20Workstation_.pdf)

Educopia Institute, OSSArcFlow: Investigating, Synchronizing, and Modeling a Range of Archival Workflows for Born-Digital Content <https://educopia.org/ossarcflow/>

# Resource page: digitization process

Vendors: Digitizing Video for Long-Term Preservation: An RFP Guide and Template

[https://guides.nyu.edu/ld.php?content\\_id=24817650](https://guides.nyu.edu/ld.php?content_id=24817650)

In-house: Minimum viable digitization station: <http://bit.ly/mindigit>

Video playback refurbishment vendors:

[https://cool.culturalheritage.org/videopreservation/dig\\_mig/repair\\_historicVTR.html](https://cool.culturalheritage.org/videopreservation/dig_mig/repair_historicVTR.html)

<http://www.zinvtrworks.com/about-ken-zin-principal-engineer-2/>

Video playback deck vendors (other than eBay or other local sources), can include warranty:

Southern Advantage: <https://www.southernadvantage.com/>

Broadcast Store: <https://www.broadcaststore.com/>

Questions so far



15 min break

# Post-digitization or born digital:

## File format validation and file structure documentation

Validation / file format identification tools:

- Audiovisual:
  - Mediainfo <https://mediaarea.net/en/MediaInfo>
  - Jhove <https://jhove.openpreservation.org/>
  - FFprobe <https://ffmpeg.org/ffprobe.html>
- Static media:
  - Exiftool <https://exiftool.org/>
  - DROID: <https://digital-preservation.github.io/droid/> (Uses PRONOM file format signatures: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)
  - Siegfried <https://www.itforarchivists.com/siegfried/> (Uses PRONOM file format signatures: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)
  - Fido: <https://openpreservation.org/products/fido/> (Uses PRONOM file format signatures: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>)

File and directory structure documentation:

- Fiwalk (part of Bitcurator environment, creates DFXML; also SleuthKit)
  - DFXML schema: <https://github.com/simsong/dfxml>
  - Filesystem is named - e.g. Fat32 or HFS or exFAT.
- Bulk extractor:
  - <https://archive.ph/2lyXm> (info) [https://downloads.digitalcorpora.org/downloads/bulk\\_extractor/](https://downloads.digitalcorpora.org/downloads/bulk_extractor/) (download)

# Preparing policy/workflow: Developing SIP, AIP, DIP

- Sometimes dictated by software, work with vendor to analyze if the software meets your needs before adoption, if you can.
- Or, you can work on a policy that frames needs and choose software from there.
- Or, you can handle it in a more DIY manner and work with micro-services or scripts in order to create a meaningful workflow.

Resources re: micro-services / scripts:

- Look into digital-curation, Code4Lib or AMIA listserv for recommendations for scripts, people on Github who write scripts, etc. Also may find consultants this way.
- Dinah Handel, Media micro-services and archival workflows at CUNY Television <https://ndsr.nycdigital.org/media-miscro-services-and-archival-workflows-at-cuny-television/>

# Resource page: digital preservation policy examples

Cornell University Library (2004):

<https://ecommons.cornell.edu/xmlui/bitstream/handle/1813/11230/cul-dp-framework.pdf?sequence=1>

ISPCR, University of Michigan:

<https://www.icpsr.umich.edu/web/pages/datamanagement/preservation/policies/index.html>

University of Washington Libraries:

[https://www.lib.washington.edu/preservation/preservation\\_services/digitization-and-digital-preservation/digital-preservation-policy](https://www.lib.washington.edu/preservation/preservation_services/digitization-and-digital-preservation/digital-preservation-policy)

# Exercise: create checksum

Please chat to us if you are on a Windows or a Mac computer.

Breakout room 1, Instructions for Mac OS:

<https://www.mjdtech.net/how-to-check-md5-checksum-in-os-x-terminal/>

Breakout room 2, Instructions for Windows OS:

<https://onthefencedevelopment.com/2017/08/15/windows-10-builtin-md5-checksum-calculator/>

# Day 2

- NDSA Levels discussion
- OAIS discussion
- Storage discussion
- Software
- Metadata
- Consultations
- Topics day 2: NDSA, OAIS, metadata, storage, software, consultations
- Exercises day 2: Discussion, Bagit demo

# NDSA Levels of Digital Preservation

Functional Area	Level			
	Level 1 (Know your content)	Level 2 (Protect your content)	Level 3 (Monitor your content)	Level 4 (Sustain)
Storage	Have two complete copies in separate locations	Have three complete copies with at least one copy in a separate geographic location	Have at least one copy in a geographic location with a different disaster threat than the other copies	Have at least three geographic locations with different disaster threats
	Document all storage media where content is stored	Document storage media indicating the resources and dependencies they require to function	Track the obsolescence of storage and media	Maximize storage and avoid single point of failure

Levels v2: <https://ndsa.org/publications/levels-of-digital-preservation/>

Working definitions of terms:

<https://mfr.osf.io/render?url=https://osf.io/rnymf/?direct%26mode=render%26action=download%26mode=render>

**TONIGHT:**

Please review the levels document and come to class tomorrow with your best understanding of where your institution is at in the grid.

# OAIS, or Open Archival Information System & TDR certification

DPC introduction to OAIS: <https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>

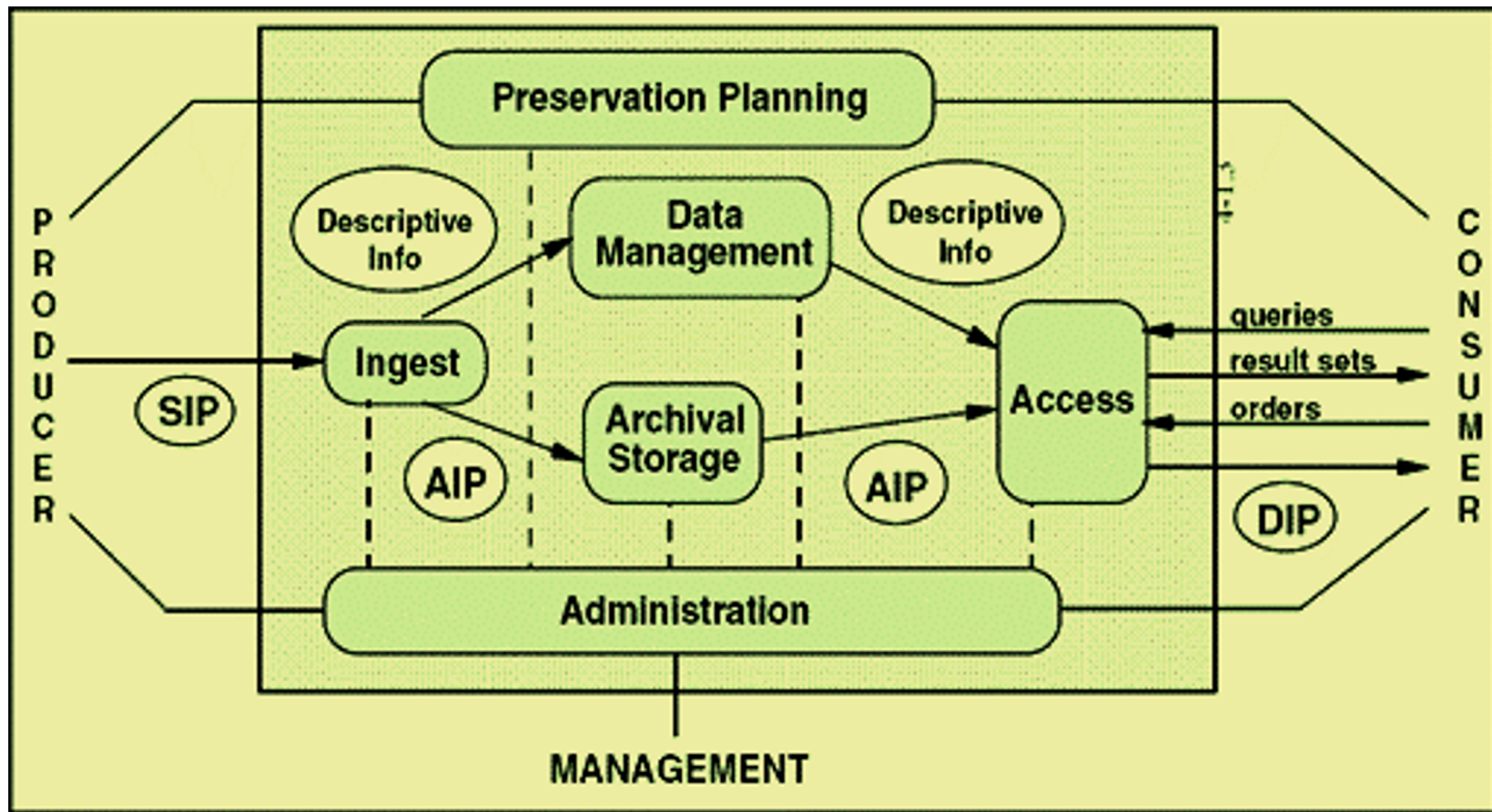
Cal Lee, Open Archival Information System (OAIS) Reference Model: <https://ils.unc.edu/callee/p4020-lee.pdf>

Rhiannon S. Bettivia, The Power of Imaginary Users: Designated Communities in the OAIS Reference Model <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/pra2.2016.14505301038>

Consultative Committee for Space Data Systems (CCSDS), OAIS Reference Model (2012): [https://digital.library.unt.edu/ark:/67531/metadc123535/m2/1/high\\_res\\_d/650x0m2.pdf](https://digital.library.unt.edu/ark:/67531/metadc123535/m2/1/high_res_d/650x0m2.pdf)

TDR certification information / TRAC Checklist / ISO 16363: <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/iso16363>





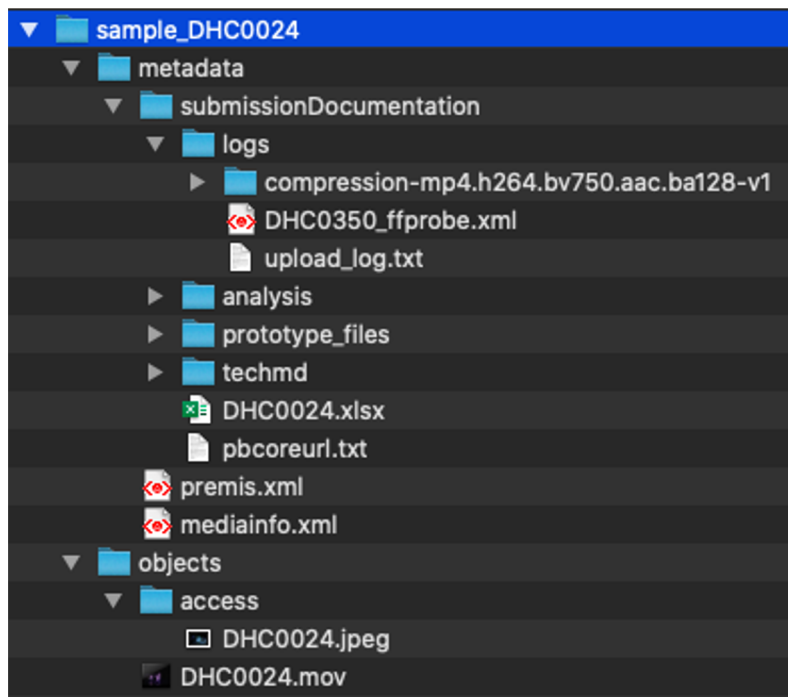
OAIS Reference Model, image courtesy Wikimedia Commons

# OAIS, or Open Archival Information System

Example AIP structure:

<https://www.archivematica.org/en/docs/archivematica-1.12/user-manual/archival-storage/aip-structure/#aip-structure>

**Example SIP structure:**



# Storage media

- Cloud storage
- Server-attached storage
- External hard drives
- LTO or other type of data tape

## Resource page: storage

IASA Technical Committee, Handling and Storage of Audio and Video Carriers  
<https://www.iasa-web.org/tc05/handling-storage-audio-video-carriers>

AVP, Cloud storage vendor profiles <https://www.weareavp.com/cloud-storage-vendor-profiles-2/>

PASIG, Preservation Storage Criteria: <http://go.ucsd.edu/2xSLGQ6>; Explanation of criteria: [https://pasig.figshare.com/articles/presentation/Acid-free\\_AIPS\\_Digital\\_preservation\\_storage\\_criteria/5415145](https://pasig.figshare.com/articles/presentation/Acid-free_AIPS_Digital_preservation_storage_criteria/5415145)

# Resource page: Software for digital preservation

- Fedora: <https://duraspace.org/fedora/resources/publications/fedora-digital-preservation/>
- Archivematica: <https://www.archivematica.org/en/>
- DSpace: <https://duraspace.org/dspace/>
- Preservica: <https://preservica.com/>
- Samvera (fka Hydra): <http://samvera.org>

## **The Collection Management System Collection:**

[https://docs.google.com/spreadsheets/d/1cXOug3qM0pNNeD\\_wssiVEv9c0W1Y5I1VDTnSPTk7fb4/edit#gid=0](https://docs.google.com/spreadsheets/d/1cXOug3qM0pNNeD_wssiVEv9c0W1Y5I1VDTnSPTk7fb4/edit#gid=0)

# Preservation metadata, digital and moving image

PBCore: Public Broadcasting Core (maintained by GBH Media Library & Archives)

- Instantiations, different physical or digital versions of one descriptive unit

PREMIS: Preservation Metadata Implementation Standard (maintained by LOC)

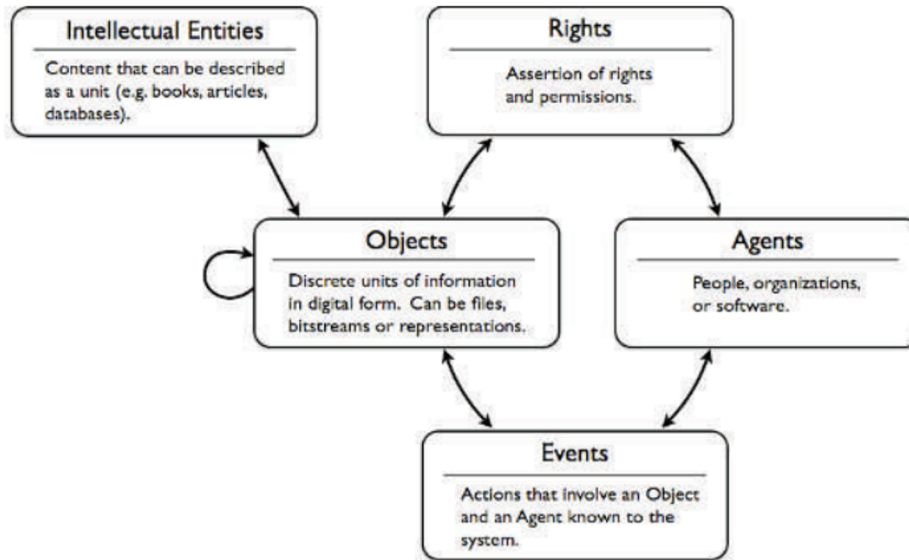
- Objects, Rights, Events, Agents
- Events: <https://www.loc.gov/standards/premis/v3/preservation-events.pdf>

METS: Metadata Encoding and Transmission Standard (maintained by LOC)

- Wraps descriptive and technical information

# Understanding preservation metadata

## 3. PREMIS DATA MODEL



# Understanding preservation metadata

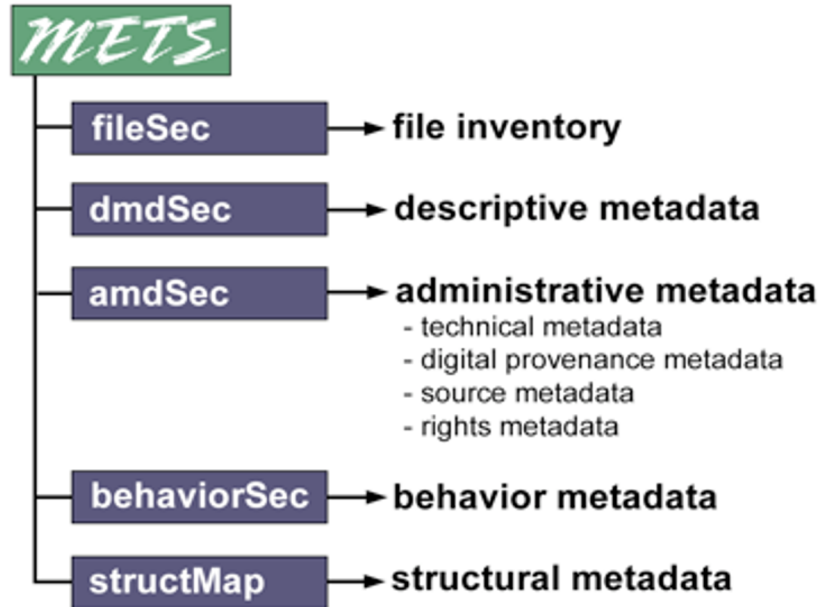


Image: <http://www.dlib.org/dlib/july08/guenther/07guenther.html>



# Resource page: metadata, digital and moving image

PBCore spreadsheet templates: <https://pbcore.org/spreadsheet-templates>

PBCore tutorial: <https://pbcore.org/tutorials> PBCore 2.1 schema: <https://pbcore.org/xsd>

PREMIS tutorial: <https://www.loc.gov/standards/premis/tutorials.html> PREMIS schema: <https://www.loc.gov/standards/premis/index.html>

Digital Preservation Coalition, recent PREMIS blog post: <https://www.dpconline.org/blog/wdpd/blog-premis-wdpd>

METS tutorial: <https://www.loc.gov/standards/mets/METSOverview.html> METS schema: <https://www.loc.gov/standards/mets/>

Dublin Core: <https://dublincore.org/>

Using PREMIS with METS: <http://www.dlib.org/dlib/july08/guenther/07guenther.html>

PREMIS Events: <https://www.loc.gov/standards/premis/v3/preservation-events.pdf>