

2-10-2016

# METS Schema: a general review

Tammy Troup

*Bucknell University*, [tlt014@bucknell.edu](mailto:tlt014@bucknell.edu)

Follow this and additional works at: [https://digitalcommons.bucknell.edu/lit\\_pubs](https://digitalcommons.bucknell.edu/lit_pubs)



Part of the [Cataloging and Metadata Commons](#)

---

## Recommended Citation

Troup, Tammy, "METS Schema: a general review" (2016). *Library and Information Technology Publications*. 9.  
[https://digitalcommons.bucknell.edu/lit\\_pubs/9](https://digitalcommons.bucknell.edu/lit_pubs/9)

This Conference Proceeding is brought to you for free and open access by the Library and Information Technology at Bucknell Digital Commons. It has been accepted for inclusion in Library and Information Technology Publications by an authorized administrator of Bucknell Digital Commons. For more information, please contact [dcadmin@bucknell.edu](mailto:dcadmin@bucknell.edu).



# METS SCHEMA

A General Review

# Managing Materials

## Print/Analog



15<sup>th</sup> century Chained Book,  
Malatestiana Library, Cesena, Italy

## Digital

```
Administrator: C:\Windows\system32\cmd.exe
L:\Library\Books>dir
Volume in drive L is MHSData
Volume Serial Number is 2064-7364

Directory of L:\Library\Books

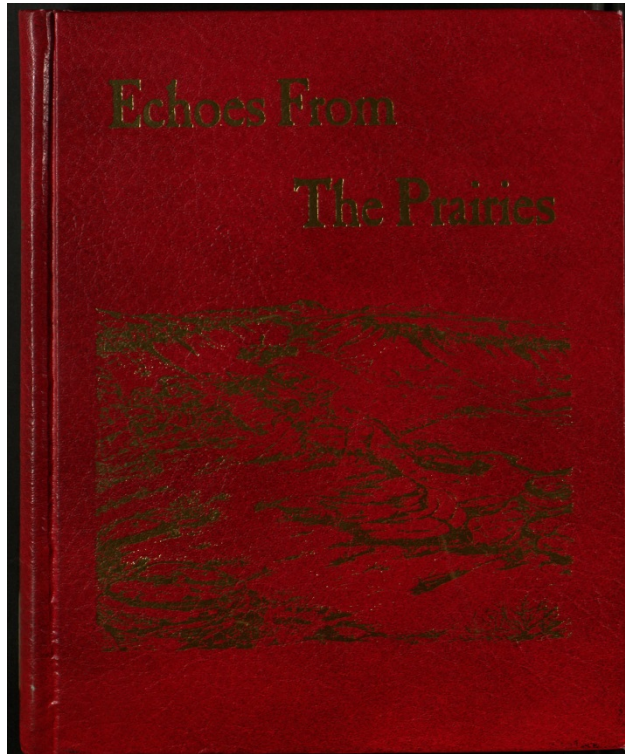
02/02/2016 03:34 PM <DIR> .
02/02/2016 03:34 PM <DIR> ..
11/17/2015 09:35 AM <DIR> 355 SU26H
11/17/2015 09:37 AM <DIR> 366.63 RM76
11/17/2015 09:37 AM <DIR> 367.9786 V33H
11/17/2015 09:38 AM <DIR> 374.2 M76
11/17/2015 09:39 AM <DIR> 379 H36B
11/17/2015 09:39 AM <DIR> 379 H36R
01/27/2016 05:24 PM <DIR> 385 I3
11/17/2015 09:41 AM <DIR> 636.08128 M76 1890
11/17/2015 09:41 AM <DIR> 636.08128 M76 1894e2
11/17/2015 09:42 AM <DIR> 636.08128 M76 1899e1
11/17/2015 09:42 AM <DIR> 636.08128 M76 1903e2
11/17/2015 09:43 AM <DIR> 636.08128 M76 1910e1
11/17/2015 09:44 AM <DIR> 658.8025 C51
11/17/2015 09:45 AM <DIR> 978.6615 h 363 h 1890 c1
11/17/2015 09:46 AM <DIR> 978.6
11/17/2015 10:09 AM <DIR> 978.6 F63 1960
11/17/2015 10:10 AM <DIR> 978.66 I73
11/17/2015 10:10 AM <DIR> 978.6611 G79L
11/17/2015 10:11 AM <DIR> 978.6615 MS3U
01/29/2016 01:16 PM <DIR> B W838n Mr. White
11/17/2015 10:12 AM <DIR> Capitol Capsules
11/17/2015 10:12 AM <DIR> CheckSums
12/21/2015 09:07 AM 235 ObjectInfo.csv
11/17/2015 10:13 AM <DIR> REF 071.9786 M762ULM
1 File(s) 235 bytes
25 Dir(s) 19,995,294,744,576 bytes free

L:\Library\Books>
```

“Books Directory ” L: drive, MHS,  
Helena, Montana

# Clues – physical (carrier)

## Book

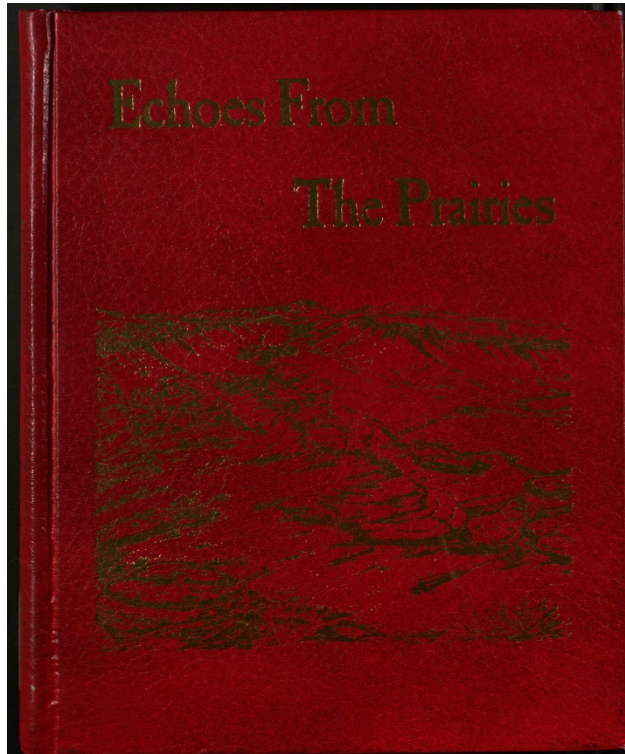


## Carrier Information

- Binding
- Paper
- Signatures or Sections
- Folios

# Clues – structural (carrier)

## Book

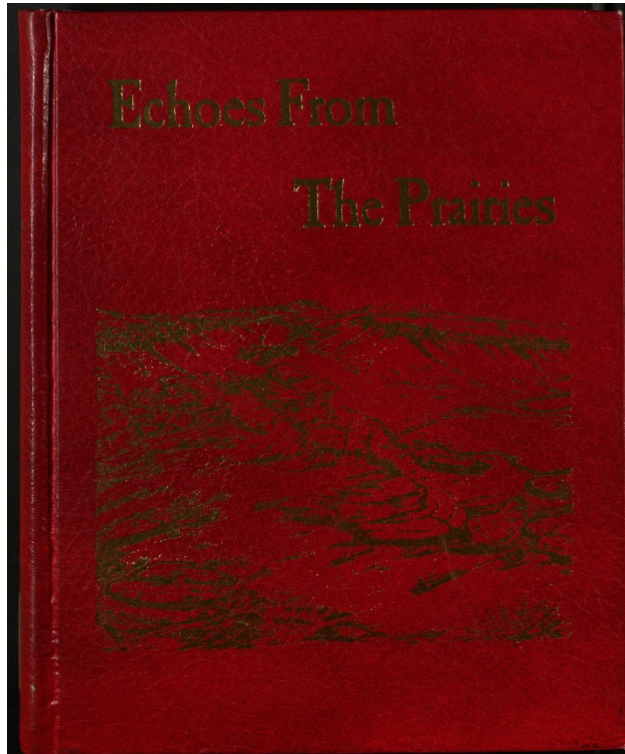


## Content Information

- Language
- Layout
- Images
- Font, point size, special characters
- Arrangement

# Clues – descriptive (transmission)

## Book

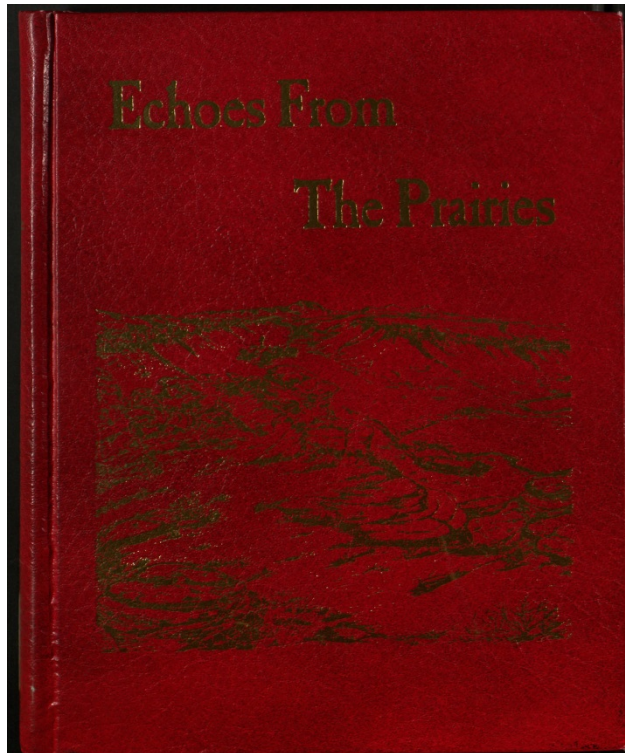


## Content Information

- Title
- Description (content)
- Culture
- Date (created)
- Publisher
- Contributor

# Clues – administrative (transmission)

## Book



## Provenance Information

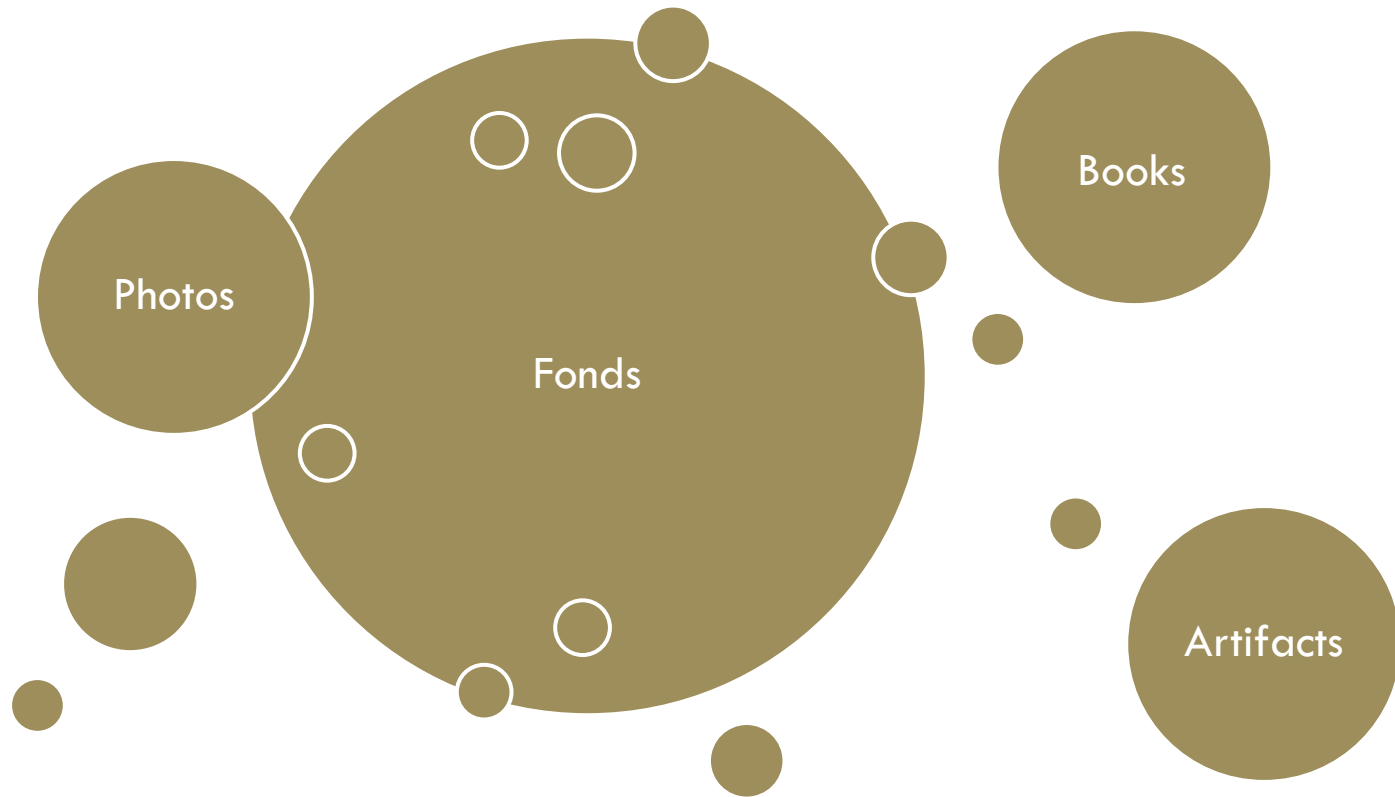
- Accession
- Creator
- Rights
- License
- Changes
- Dates (events)
- Contributor (changes)



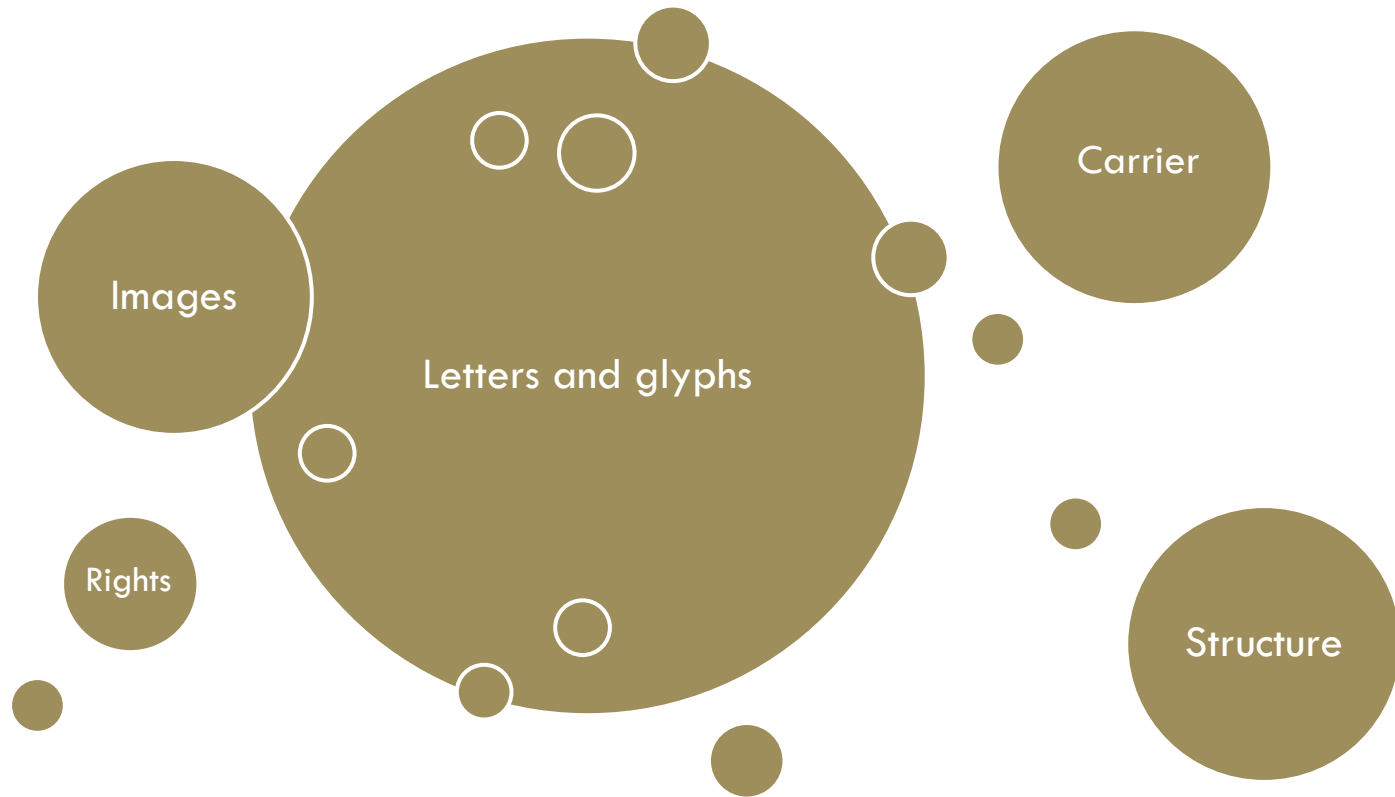




# Limits of Discovery Records



# Limits of Discovery Records





One Record to Rule  
Them All...

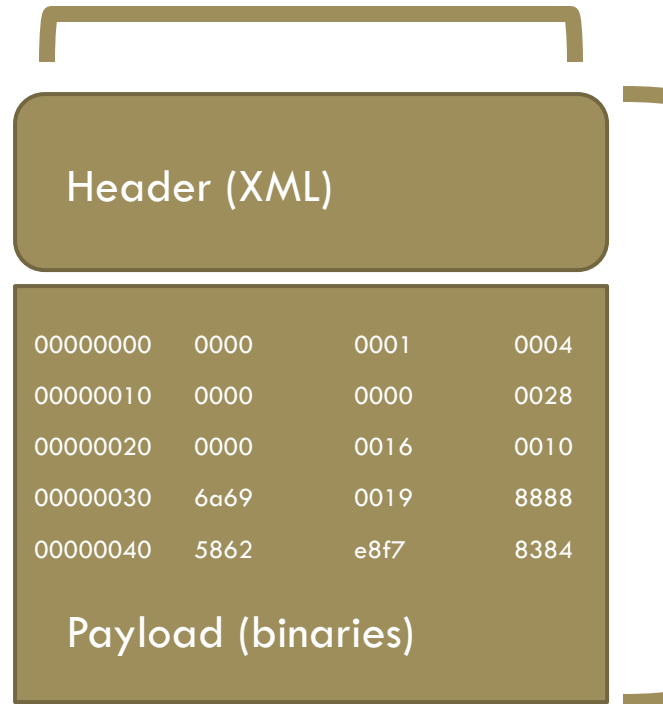


...sorta

# Digital Files

Date  
File Signatures  
Owner

Rights  
Structural Metadata  
Styles Metadata



**Digital File Container**  
(.docx -ZIP archive file with XML and binaries)



Human/Machine



Human/Machine

## Administrative

### Metadata

Creator, Date, Title, Related Files, Rights, Contributor, Identifiers, File Names, etc.

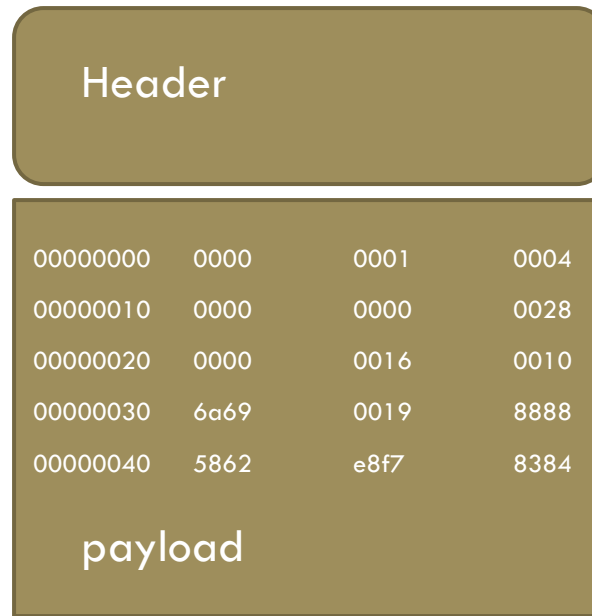
### Descriptive Metadata

Description of content, Keywords (tags), Subject-specific information, Coverage, Dates, etc.

## Technical & Preservation

### Metadata

Software, Hardware, Structure, Format, Colorspace, Size, Checksum, Events/Actions, Dates, etc. Some of this metadata is automatically captured by originating software.



## Digital File Container

(.txt, .doc, .jpg, .tiff, .png, .xlsx, .pdf)



# Digital Object

Header

00000000	0000	0001	0004
00000010	0000	0000	0028
00000020	0000	0016	0010
00000030	6a69	0019	8888
00000040	5862	e8f7	8384

payload

XML declaration

```
<METS:fileGrp>
<METS:file ID="MUSIC_FOLDER_DC" MIMETYPE="text/xml">
<METS:FContent>
<METS:xmlData>
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>Music Folder</dc:title>
<dc:creator>Chris</dc:creator>
<dc:subject>music</dc:subject>
<dc:description>This folder contains my music files.</dc:description>
</oai_dc:dc>
</METS:xmlData>
</METS:FContent>
</METS:file>
</METS:fileGrp>
```

XML CONTENT

Digital file to be preserved and XML sidecar



**txt**--manifest



**md5**--checksum



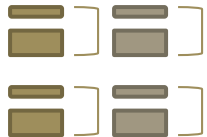
**XML**--collection level



**txt**--collection description



**pdf**--rights documents



**tiff & XML**--digital objects



**pst**--communication files

**Submission Information  
Package (SIP)**

**Ready for Preservation**





Digital Services

Archives

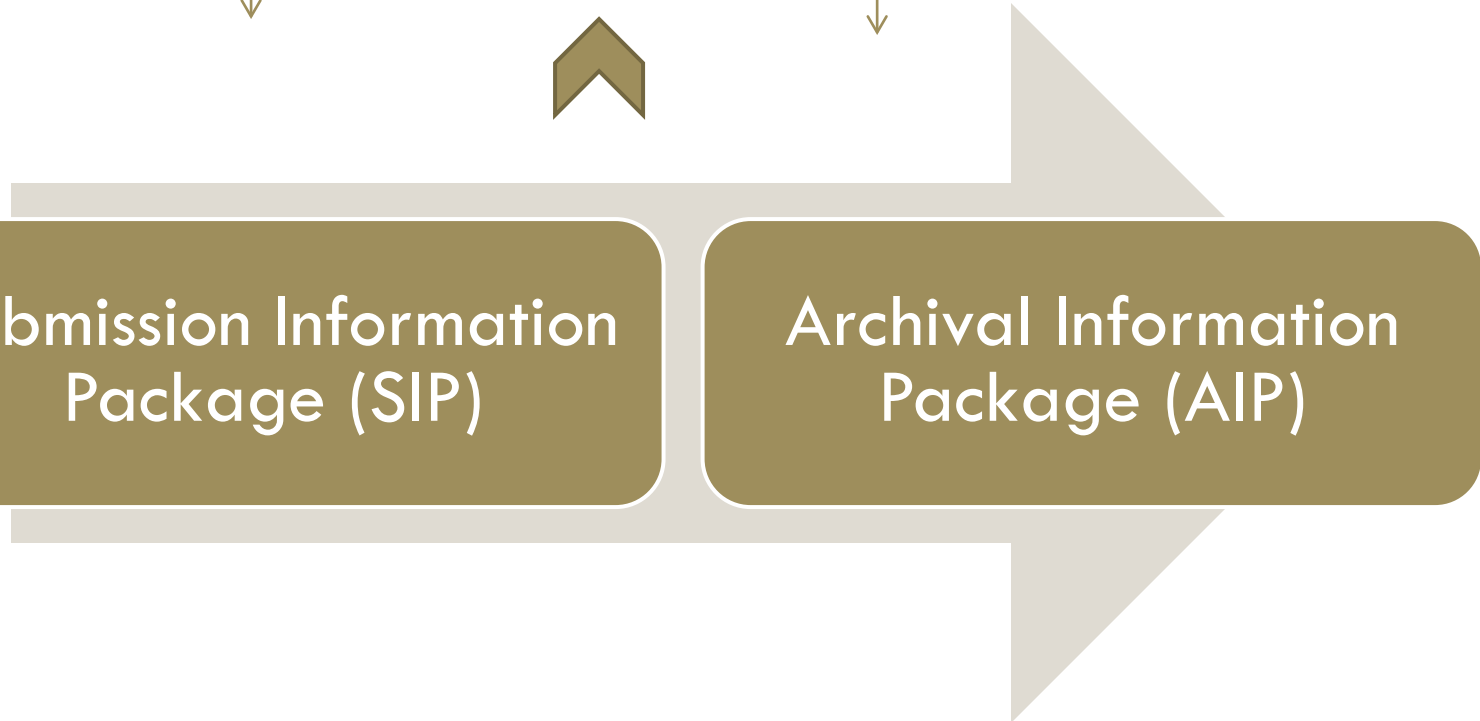


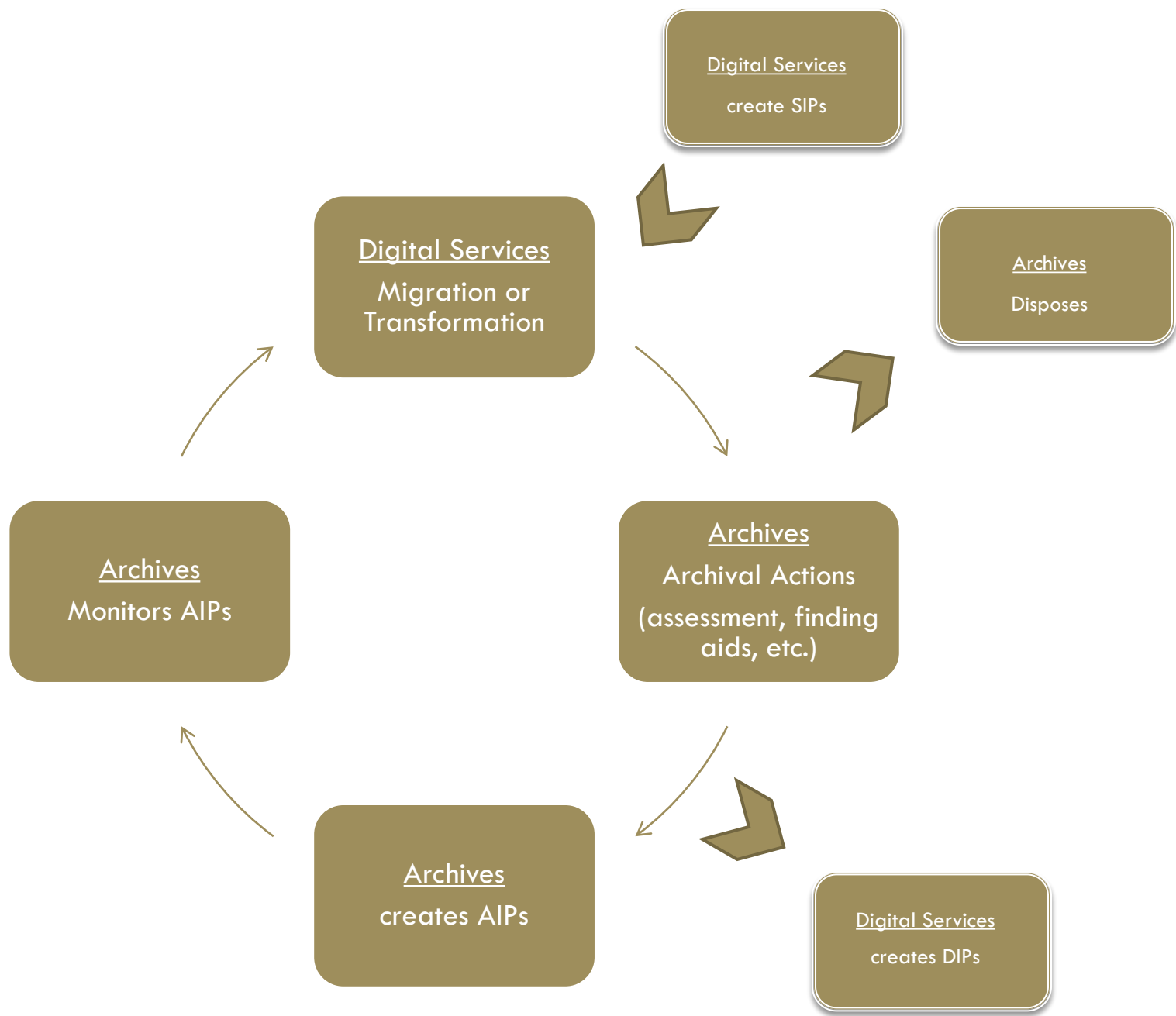
Dissemination  
Information Package  
(DIP)

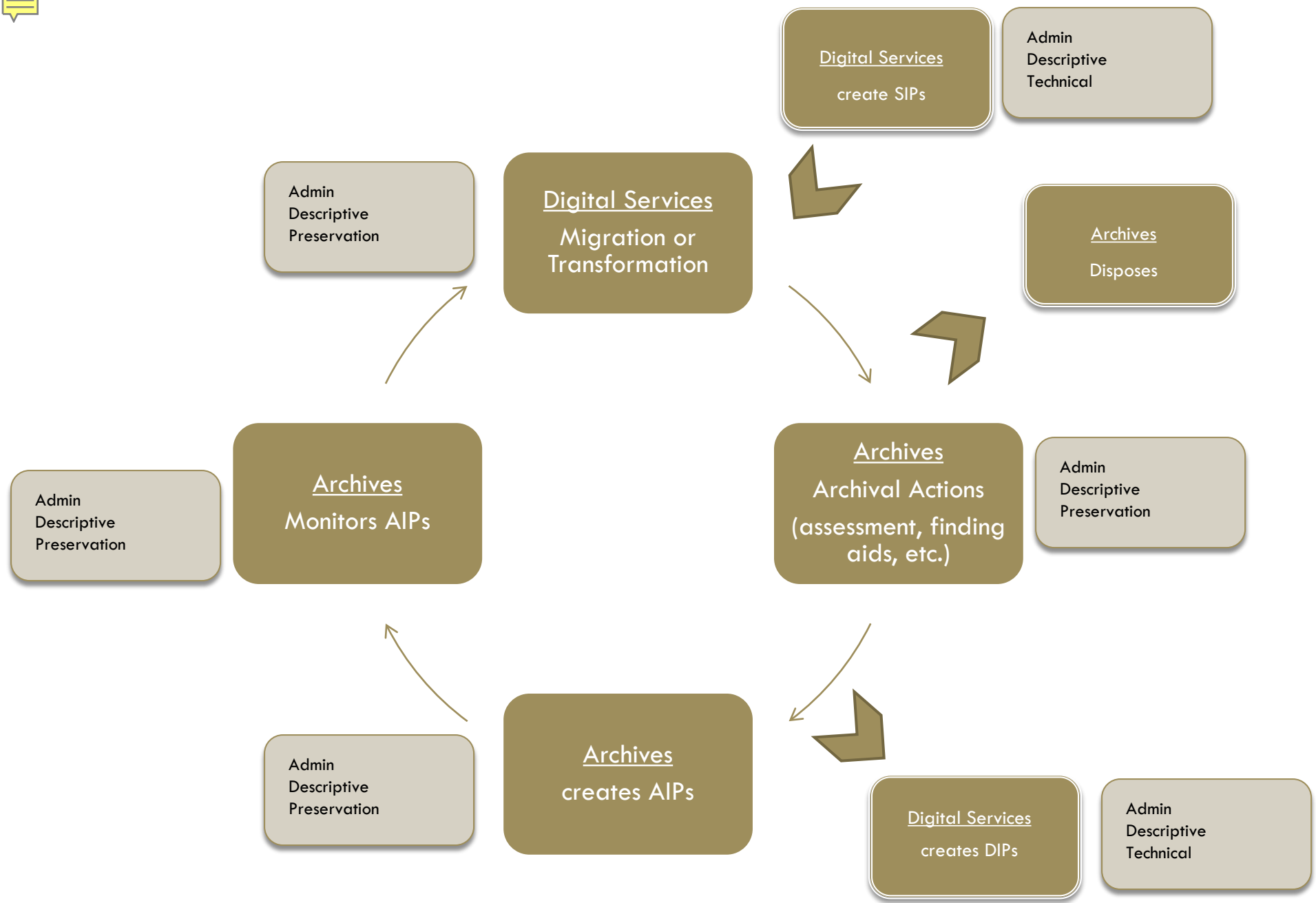


Submission Information  
Package (SIP)

Archival Information  
Package (AIP)







# Machine-generated Metadata

Microservices/ Archival Activities	Ingest	Processing	Access	Storage	Maintenance
Event Tracking	File Dedupe	Auto Metadata Harvest	Public Interface	Auto AIP Creation	Auto Recovery
Administrative Reports	Virus Scan	Auto Metadata Creation	Auto DIP Creation	Reliable, Long-term Bit Preservation	Monitoring
Authoring	Fixity Check	Rights Management	Package Metadata	Redundancy	Migration
Descriptive Metadata	Copy	Auto Unique ID	Auto SIP Creation	Geographically Disparate Storage	Exit Strategy



# Container Standards— Carry and transmit

---

- **METS** – Metadata Encoding and Transmission Standard, schema developed by LOC
- **MARC** – Machine Readable Cataloging
- **MPEG21** – Moving Pictures Experts Group, open framework for multimedia applications
- **OAI-ORE** – Open Archives Initiative Object Reuse and Exchange
- **MXF** – Material Exchange Format
- **AVI** – Audio Video Interleaved



# Content rules and schemas

---

- ❑ AACR2 – Anglo-American Cataloguing Rules (2)
- ❑ RDA – Resource Description and Access
- ❑ ALTO—Analyzed Layout and Text Object, schema developed by the LOC
- ❑ Dublin Core—Dublin Core Metadata Initiative
- ❑ MODS—Metadata Object and Description Schema, schema developed by MARC standards office and LOC
- ❑ PREMIS—Preservation Metadata developed by LOC
- ❑ And others...



# LOD & Registries

---

- ❑ PRONOM <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>
- ❑ Unified Digital Format Registry (UDFR) <http://www.udfr.org/>
- ❑ Planets Core Registry  
[http://justsolve.archiveteam.org/wiki/Planets\\_Core\\_Registry](http://justsolve.archiveteam.org/wiki/Planets_Core_Registry)
- ❑ Digital Record and Object Identification (DROID) <http://digital-preservation.github.io/droid/>
- ❑ Linked Data Registry  
<http://ukgovld.github.io/ukgovldwg/guides/registry.html>
- ❑ ORCID <http://orcid.org/>
- ❑ Getty Vocabularies as Linked Open Data  
<http://www.getty.edu/research/tools/vocabularies/lod/>



# What is METS

(©LYRASIS, presented for educational purposes only, not for distribution)

- METS is an XML Schema designed for the purpose of creating XML document instances that express the hierarchical structure of digital library objects, the names and locations of the files that comprise those objects, and the associated metadata.
- A package for
  - ▣ Descriptive and administrative metadata (technical, digital provenance, rights, source)
  - ▣ Structural metadata
- Focus is on complex digital library objects



# More About METS

(©LYRASIS, presented for educational purposes only, not for distribution)

- Records relationships among the metadata and among the pieces of the complex objects
- Can be a unit of storage (e.g. OAIS AIP) or a transmission format (e.g. OAIS SIP or DIP)
- Content-type independent
- Batch processing for creation, processing, retrieval, and presentation
- Can be encoded with a text editor, XML editor, or a forms-based user interface

# More Provision of METS

(©LYRASIS, presented for educational purposes only, not for distribution)

- Identifying the files or parts of files that comprise the content of a digital entity, and expressing their structure
- Linking Descriptive metadata with digital content
- Linking Administrative metadata with digital content
- Linking behavior definitions and program code with digital content and with associated metadata
- Wrapping digital content, and associated metadata



# METS sections





# Example (1) METS/ALTO at MHS

---

- Title METS Object (example)
- Issue METS Object (example)
- Reel METS Object (example)



# Example (2) LC METS

---

- <http://www.loc.gov/standards/premis/louis-2-1-mets.xml>
- <https://www.loc.gov/item/gottlieb.09611/>

# Example (3) Linking Ideas (metadata)



# Feeling a little underwater?



# Need to know...

---

- Understand what an information package is...
- Understand how to basically describe an improperly bundled information package and a properly bundled information package...
- How and when to embed metadata in files...
- How and when to capture additional metadata...
- How and why to advocate for DAM...
- How and why to advocate for Digital Preservation...



# Need to create...

- MHS Core – core METS schema for digital object item types and content display; container standards need identified and content standards need identified.
  - ▣ Montana Memory Project application profile (& DPLA)
  - ▣ ArchivesWest application profile
  - ▣ SirsiDynix application profile
  - ▣ NDNP application profile
- Temporary workflows to capture retrospective and ongoing project information
- Permanent solution involving DAM and Digital Preservation systems

# Citations

- Slide 1—Title
- Slide 2 – John Willis Clark - <http://www.ib.hu-berlin.de/~wumsta/Milkau/109-2.jpg>: Fritz-Milkau-Dia-Sammlung, erstellt in der Photographischen Werkstatt der Preußischen Staatsbibliothek von 1926-1933 originally from The Care of Books by John Willis Clark (Fig 94, p132), Public Domain, <https://commons.wikimedia.org/w/index.php?curid=135286>
- Slide 20, 21, 22 – Guenther, Rebecca. “Metadata for Digitization and Preservation. Part 2: METS” LYRASIS training 4 February 2016. Slides 3, 5, 9.
- Slide 29 – fieldsbh. “Playing with a ball of yarn.” Flickr. Creative Commons license. 20070402. <https://www.flickr.com/photos/fieldsphotos/445629313>
- Slide 30 – Nyveen, Jeff. “Mermaids in Montana.” No Regrets Blog. 20100916. Educational use only, not for distribution. <http://travel.jeffnyveen.com/category/places/usa/devils-tower-national-monument/>

Q/A?