

#### University of Nebraska at Omaha DigitalCommons@UNO

Criss Library Faculty Proceedings & Presentations

Dr. C.C. and Mabel L. Criss Library

10-5-2018

#### **Best Practices for Digital Collections**

Corinne Jacox Creighton University, corinnejacox@creighton.edu

Yumi Ohira University of Nebraska at Omaha, yohira@unomaha.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/crisslibfacproc

Part of the Library and Information Science Commons

Please take our feedback survey at: https://unomaha.az1.gualtrics.com/jfe/form/ SV\_8cchtFmpDyGfBLE

#### **Recommended Citation**

Jacox, Corinne and Ohira, Yumi, "Best Practices for Digital Collections" (2018). Criss Library Faculty Proceedings & Presentations. 93.

https://digitalcommons.unomaha.edu/crisslibfacproc/93

This Presentation is brought to you for free and open access by the Dr. C.C. and Mabel L. Criss Library at DigitalCommons@UNO. It has been accepted for inclusion in Criss Library Faculty Proceedings & Presentations by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.





# **Best Practices for Digital Collections**

Corinne Jacox, Catalog/Reference Librarian, Creighton University Law Library Yumi Ohira, Digital Initiatives Librarian, University of Nebraska at Omaha Libraries

2018 NLA/NSLA Joint Conference, October 5th, 2018

### **Planning**

#### Create New Research Opportunities for Users

- Available to a variety of users
- Enhance understanding of the digital objects

#### Planning is Key

- Institution's mission and goals
- Costs and capabilities for long term maintenance
- Standards -- documentation, formats, metadata
- Timeline
- Define the scope (characteristics of objects to be digitized)
  - How many, formats, sizes, unusual characteristics, condition
- Workflow
- Equipment -- optimize quality and level of production



#### **Policies**

Access

Content

Committee

Copyright

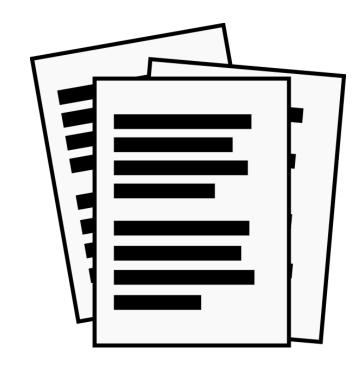
Metadata

Preservation

Withdrawal

#### **Examples**

<u>University of Chicago Library</u> <u>University of Hawai'i at Mānoa Library</u>



#### Institutional Repository Policies

This page overviews policies that guidig Knowledge@UChicago, the University of Chicago's institutional repository.



#### Scope

#### Contributors to Repository

Contributors must be affiliated with the University of Chicago and have valid CNetID credentials to deposit work in the repository. Alumni and other affiliated users without CNet credentials should contact the Knowledge@UChicago team.

#### Scope of Deposited Content

- . Deposited content can be scholarly, creative, research-related, or teaching resources.
- . Deposited content should be in a completed state, rather than in-progress and regularly updated.
- . Contributors may deposit content created prior to joining the University of Chicago.
- . Contributors must be willing and able to grant the University the non-exclusive rights to both preserve and make their work available through Knowledge@UChicago.
- · Deposited content must be in a digital format.
- If the deposited content is part of a series, other works in that series should also be deposited when possible so that we can offer a full and complete collection.

#### **Supported Content Types**

The following content types are among those that will be accepted:

- · publications (including preprints)
- dissertations
- · masters and undergraduate theses
- · working papers and technical reports
- white papers
- . datasets up to 1GB (data sets must be complete and ready to for use, and must include a readme file. At this time, no classified/restricted/confidential data can be accepted.)
- · conference presentations
- · conference posters
- campus-based publications
- · course-based publications
- · teaching and learning resources
- · audio and video recordings

This is a non-exhaustive list. Contributors are welcome to deposit content that can be characterized as "scholarly, creative, research-related, or teaching resources" and to contact us with any questions about materials.



Library / Research Guides / Library Departments and Services / Digital Collections and Repository Program (D-CARP) / Prioritization and Selection Criteria

### Digital Collections and Repository Program (D-CARP): Prioritization and Selection Criteria

Search this Guide Search

Home Policies & Procedures → Background Digital Collections & Repositories Grants → Related Services Initiatives

#### PRIORITIZATION AND SELECTION CRITERIA

#### Metadata information - Cataloging/Digital Library Program Task Force Requirements

- Digitization Project Survey Information needed about project scope, etc.
- . Digitization Project Survey Online fillable
- Metadata information Cataloging/Digital Library Program Task
  Force Requirements

#### UHM Library Project Selection Factors

Project Evaluation Criteria

Policy Considerations for Digital Projects - Hawaiian/Pacific Collections

UHM Library holds a large number of collections that would be appropriate for digitization and online access. However, digitization projects are costly and require a commitment of staff time. The following list of criteria is recommended to guide selection of collections of analog materials for conversion to digital format. Selection is an activity led by content managers and specialists with the help of D-CARP.

#### Copyright Status

What is the copyright status of the materials? Most material considered for digitization and access on the open Web falls into one of the following three categories:

- · Public domain: works that never were, or are no longer covered by copyright. Works in the public domain may be used without permission.
- All works published before January 1, 1923.
- . Works published between 1923 and 1964 and not renewed in the 28th year.
- Works published without copyright notice before 1989.
- Unpublished works whose author died before 1932; otherwise, the term is life plus 70 years.
- . Works for which the copyright is held by the University
- . Works for which we have secured permission to digitize

### **Staffing**

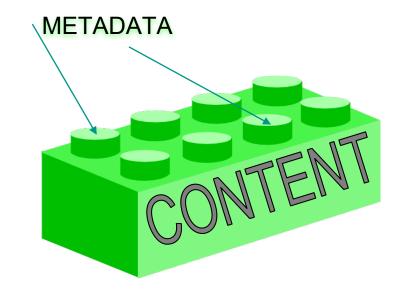
- Project Manager -- oversees daily operations and maintains the budget, timeline, and workflow
- Collections Assessor -- selects originals, checks their condition and makes conservation recommendations, rehouses originals as necessary, and reshelves them once digitization and cataloging are complete
- **Database Manager** -- creates and maintains databases for the project
- **Scanning Technician** -- handles original objects, creates scans and surrogates, and produces backup files on the storage media
- **Quality Control Technician** -- checks the image files generated by the scanning technician against benchmarks
- **Cataloger** -- creates or edits records for digital images of originals included in the project
- **Web Manager** -- designs and maintains the website housing the project





### Metadata

- Data about data
  - Description and context of the data
  - Helps to organize, find and understand data
- Recognized standard
  - Easier to exchange/migrate data with other systems
  - No standard can cover all materials for all users
- Three types
  - Descriptive
  - Administrative
  - Structural





## Descriptive Metadata

- Describes resource for discovery and identification
- Includes elements such as title, abstract, author, dates, subjects, keywords, and unique identifiers (PURL, handle)
- General standards
  - DCMI (Dublin Core Metadata Initiative)
  - MODS (Metadata Object Description Schedule)
  - MARC (MAchine-Readable Cataloging)
- Standards for special descriptions
  - Data Sets for Social Sciences: DDI (Data Documentation Initiative)
  - Archive: EAD (Encoded Archival Description)



Illustration by Jørgen Stamp, Digitalbevaring.dk CC Attribution 2.5 Denmark

### **Administrative Metadata**

- Provides information to help manage a resource
  - When and how it was created
  - File type
  - Technical data on creation and quality control (critical for migration and longterm sustainability of the digital resource)
  - Who can access it
- Subsets
  - Rights management metadata (intellectual property rights)
  - Preservation metadata (information needed to archive and preserve a resource)

#### **Sample Elements**

- Scanner type and model
- Resolution
- Bit depth
- Color space
- File format
- Compression
- Light source
- Owner
- Copyright date
- Copying and distribution limitations
- License information
- Preservation activities

#### Structural Metadata

- Facilitates navigation and presentation of electronic resources
  - Provides information about the internal structure of resources including page, section, chapter numbering, indexes, and table of contents
  - Describes relationship among materials (e.g., photograph B was included in manuscript A)
- Example
  - METS (Metadata Encoding & Transmission Standard)

#### **Sample Elements**

- Structuring tags
  - Title page
  - Table of contents
  - Chapters
  - Parts
  - Errata
  - Index
  - Sub-object relationship (e.g., photograph from a diary)

### **Metadata Best Practices**

- Mountain West Digital Library Dublin Core Application Profile Version 2.0 (2011)
- Lyrassis Digital Toolbox Metadata
- CDP Metadata Working Group Dublin Core Metadata Best Practices Version 2.1.1
   (2006)
- Best Practices for CONTENTdm and Other OAI-PMH Compliant Repositories:
   Creating Sharable Metadata Version 3.1 (OCLC 2013)
- Orbis Cascade Alliance Dublin Core Best Practice Guidelines Version 2.3 (2018)
- Nebraska Library Commission Nebraska Memories Metadata Guidelines (2009)



#### Contents

Introduction	1
I. Best Practices for All Fields	1
Use of semicolons	1
MARC mapping	
Requirement status	
Adding local fields not covered by this profile	2
Use of the term "resource"	2
Same field, multiple vocabularies	
Additional guidelines	
II. Explanation of Table Components	3
III. Element Tables (in alphabetical order)	5
abstract	5
alternative	5
contributor	6
conversionSpecifications	7
coverage	7
creator	8
date	9
description	10
extent	10
format	
genre	
identifier	
isPartOf	
language	•
publisher	
relation	
rights	and the second s
source	
spatial	
subject	
tableOfContents	
temporal	
title	
transcription	
type	19



Element Name	creator			
Label	Creator			
DC Definition	An entity primarily responsible for making the resource.			
Is Field Required?	Mandatory if applicable			
Is Field Repeatable?	Yes			
How to Use	Person or entity responsible for creating intellectual content of resource such as a person, organization or service. Prefer form of name as verified in the Library of Congress Name Authority File (LCNAF). If name is not listed there, give name in the following format: Last name, First name, Middle initial and period, year of birth and/or death if known, separated by a hyphen.  For further help in formatting names not found in LCNAF, consult a cataloging resource such as the Anglo-American Cataloguing Rules (AACR2), Resource Description and Access (RDA), or Describing Archives: A Content Standard (DACS).			
Refines/Refinement	None			
Schemes	Library of Congress Name Authority File (LCNAF)			
DC Mapping	dcterms:creator			
MARC Mapping	100 1# (Main Entry-Personal Name), or 110 2# (Main Entry-Corporate Name), 111 1# (Main Entry-Conference Name) or 700/710/711			



# **Privacy**

Establish procedures and policies to protect the interests of donors, individuals, groups, and institutions whose public and private lives and activities are recorded in their holdings

Access restrictions on collections to ensure that privacy and confidentiality are maintained

Users' rights to privacy by maintaining the confidentiality of their research and protecting any personal information collected about the users

Society of American Archivists



Home

**BROWSE THE COLLECTION** 

DEFENDANTS

WEBSITES

**FURTHER READING** 

THOMAS RONALD DELANEY

#### SEARCH THE COLLECTION

. Search the Delaney Tokyo Papers

**Delaney Tokyo Papers** 

# **Delaney Tokyo Papers**

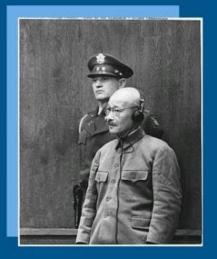


Photo credit: Tojo (left) https://www.flickr.com/photos/pingnews/441531324

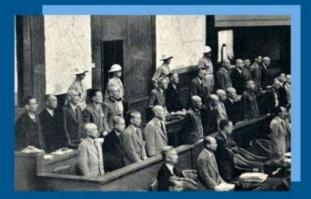


Photo credit: Defendants (right) http://modernnotion.com/verdicts-tokyo-war-crime-trial/

# Def. Doc. #2988 -- Defense Individual Summation, TOJO, Hideki, Section N-24

View/Open

TojoDefenseSummation.pdf (69.57Mb)

Author International Military Tribunal of the Far East Kiyose, Ichiro, 1884-1967

Date April 9-12, 1948

Blewett, George Francis

Subjects World War (1939-1945); Tokyo Trial (Japan: 1946-1948); War crime trials; Japan; War criminals; War crimes; Trials, litigation, etc.; Tojo, Hideki, 1884-1948

Number of Pages 347 pages

The Tokyo War Crimes Trial: Index and Guide Volume III Pages 1107-1110

The Tokyo War Crimes Trial: Transcript of the Proceedings Pages 47274-47526 Description

Individual summation on behalf of TOJO, Hideki, Prime Minister of Japan, read to the Tribunal by Blewett and Kiyose April 9-12, 1948. It contends that TOJO was not individually guilty of the charges against him, that "no act performed by him during the entire period of the indictment was criminal," and that no evidence offered by the Prosecution proved him guilty of criminal commission or criminal omission. One of the major points was that Japan was acting in self-defense.

Ichiro Kiyose, Japanese Counsel; George Francis Blewett, American Counsel

The Creighton University Law Library provides access to these materials for educational and research purposes, and for personal enjoyment. Permission to reproduce the file in any form other than for personal individual use must be obtained in writing from the Creighton University Law Library.

URI

http://hdl.handle.net/10504/74576

Collections
Delaney Tokyo T

Delaney Tokyo Trial Papers

# Def. Doc. #2988 -- Defense Individual Summation, TOJO, Hideki, Section N-24

dc.contributor.author	International Military Tribunal of the Far East
dc.contributor.author	Kiyose, Ichiro, 1884-1967
dc.contributor.author	Blewett, George Francis
dc.date.accessioned	2016-01-19T18:41:08Z
dc.date.available	2016-01-19T18:41:08Z
dc.date.issued	1948
dc.identifier.uri	http://hdl.handle.net/10504/74576
dc.description	1107-1110
dc.description.abstract	Individual summation on behalf of TOJO, Hideki, Prime Minister of Japan, read to the Tribuna I by Blewett and Kiyose April 9-12, 1948. It contends that TOJO was not individually guilty of the charges against him, that "no act performed by him during the entire period of the indictment was criminal," and that no evidence offered by the Prosecution proved him guilty of criminal commission or criminal omission. One of the major points was that Japan was acting in se If-defense. Ilchiro Kiyose, Japanese Counsel; George Francis Blewett, American Counsel
dc.description.tableofcontents	The Right of Self-Defense Who is to Decide the Existence of the Right of Self-Defense? The Right of Self-Defense and the True Nature of the China Incident Only Way Left for Jap an The Hull Note - 26 November Japan's Decision Argument of Prosecution on Right of Self-Defense Stationing of Troops in French Indo-China The Pacific War Was not Premeditated The Nine Power Treaty and the situation of Japan in 1941 The Greater East Asia policy was not of Aggressive Nature That the Second and Third KONOYE Cabinets and t

dc.rights	The Creighton University Law Library provides access to these materials for educational and research purposes, and for personal enjoyment. Permission to reproduce the file in any form other than for personal individual use must be obtained in writing from the Creighton University Law Library.					
dc.title	Def. Doc. #2988 Defense Individual Summation, TOJO, Hideki, Section N-24					
dc.title.alternative	Defense Document 2988 Section N-24					
dc.type	Other Format					
dc.rights.holder	Creighton University Law Library					
dc.date.day	April 9-12, 1948					
dc.description.note	Box 2, Folder 15					
dc.description.specialnote	47274-47526					
dc.description.pages	347 pages					
dc.subject.fast	World War (1939-1945)					
dc.subject.fast	Tokyo Trial (Japan : 1946-1948)					
dc.subject.fast	War crime trials					
dc.subject.fast	Japan					
dc.subject.fast	War criminals					
dc.subject.fast	War crimes					

Trials, litigation, etc.

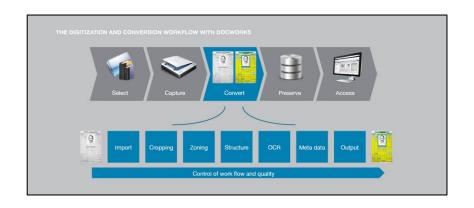
Tojo, Hideki, 1884-1948

dc.subject.fast

dc.subject.fast

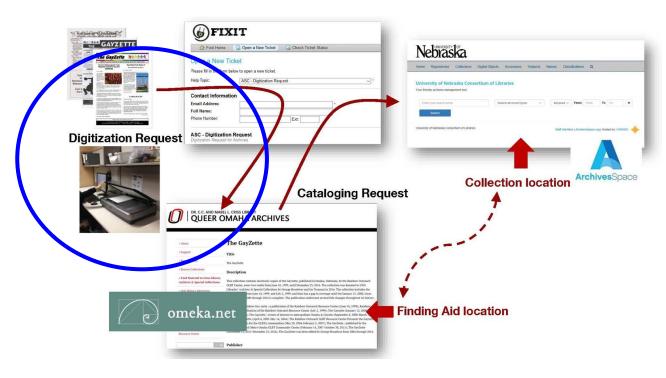
dc.subject.fast	World War (1939-1945)				
dc.subject.fast	Tokyo Trial (Japan : 1946-1948)				
dc.subject.fast	War crime trials				
dc.subject.fast	Japan				
dc.subject.fast	War criminals				
dc.subject.fast	War crimes				
dc.subject.fast	Trials, litigation, etc.				
dc.subject.fast	Tojo, Hideki, 1884-1948				
dc.url.fast	http://id.worldcat.org/fast/1180924				
dc.url.fast	http://id.worldcat.org/fast/1709967				
dc.url.fast	http://id.worldcat.org/fast/1170459				
dc.url.fast	http://id.worldcat.org/fast/1204082				
dc.url.fast	http://id.worldcat.org/fast/1170469				
dc.url.fast	http://id.worldcat.org/fast/1170465				
dc.url.fast	http://id.worldcat.org/fast/4277				
dc.url.fast	http://id.worldcat.org/fast/1423712				
dc.conversion.creation	application/PDFThis 69,500,000 byte file is derived from a 300 dpi uncompressed PDF file that was scanned from the original using a Xerox WorkCentre 7120, color configuration. ABB YY FineReader (Build 10.0.104.206; Part #700/28) was used for text recognition.				
dc.date.digital	2015-12-18				
dc.identifier.viaf	http://viaf.org/viaf/70380561				

### **Digitization Workflow**



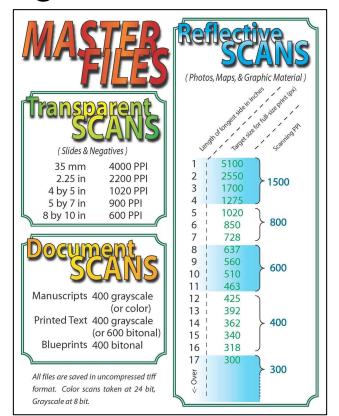
- (1) Digitization Request
- (2)Capture (Scanning process)
- (3)Convert (Edit and save digital objects)
- (4)Preserve
- (5)Access

### Digitization Workflow - (1) Digitization Request



### **Digitization Workflow**

## - (2) Capture



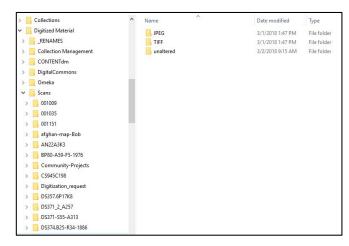








- Unaltered TIFF (Master Image)
- Edited TIFF (Dilivertive Image)
- Edited JPEG (Access Image)





### Digitization Workflow

I. Institutional name	Example 1		Example 2		Examp	le 3				
A.Parent Organization	kc	Kansas Cosmosphere	fhsu	Fort Hays St University	I the	fhsu		fhsu Fort Hays St Universit		,
B. Organization			fl	Forsyth Libr	ary gs	gs		ate School		
C. Department			ar	Archives & Sp Collection		al				
II. Object ID										
A. Collection	as204_	Apollo 1 Investigation	_wkohp_	Western Kar Oral Histor Project		th		collection		
B. Box within Collection	ecs_	Environmental Control System								
C. Folder within Box	f[x]	Folder Number x								
D. Document within Folder	d[x]	Document Number x								
F. Lastname-firstname of author / subject			ryan-stella_	Stella Ryan ( Interviewe	' I tacha-d	tacha-dustin		Dustin Tacha Author		
III. Part Designator	V. Version of Object	t								
A. Part Within Document (single)	A. Optimized file	0	(TIFF & JPEG) File optimized for use			0		(TIFF & JPEG) optimized for		
B. Reverse of Document  IV. Page Designator	B. Web-ready file	w	(JPEG)File altered for web-delivery	w	(MP3) File altered for web-delivery			(JPEG)File alte		
A. Pages within	VI. File Format									
document (Combination PDF)	A. File Extension	.jpg / .tiff / .pdf	File format used	.jpg / .tiff /.pdf / .wav / .mp3	File format used	format usedipg/.tiff/.pdf File format		File format us		
1011	Example 1: kcas204_e Example 1: kcas204_e				Single Cosmosphere document (JPEG Web-use) Combination Cosmosphere PDF document					
		vkohp_ryan-stellaow.n vkohp_ryan-stella_p1-	Oral History for WKOHP (MP3 Web-use file) Combination Oral History Transcript PDF document							
	Example 3: fhsugsth_tacha-dustin_p1-100.pdf Example 3: fhsugsth_tacha-dustin_p1.tif				Combination Thesi Single Thesis docu					

## - (3) Convert

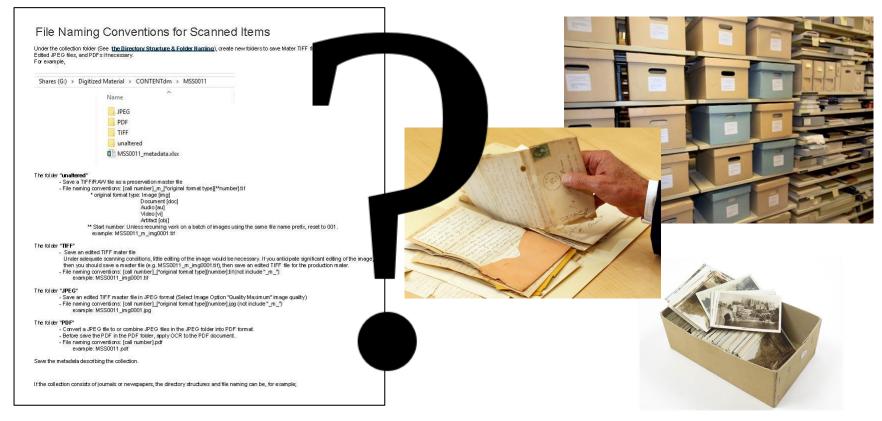
#### **File Naming Conventions:**

kcas204\_ecs\_f1d1p1ow.jpg kcas204\_ecs\_f1d1-d100.pdf



fhsuflar\_wkohp\_ryan-stellaow.mp3 fhsuflar\_wkohp\_ryan-stella\_p1-100.pdf

fhsugsth\_tacha-dustin\_p1-100.pdf fhsugsth\_tacha-dustin\_p1.tif





MSS0034\_1978\_Wether\_m\_001.tiff MSS0034\_1978\_Wether\_001.jpg

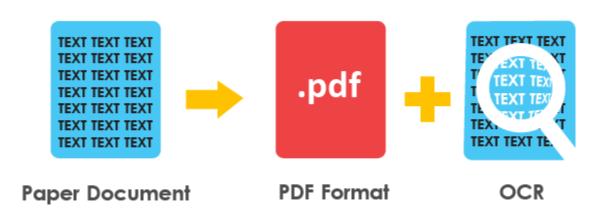
UNO0012\_Boys\_of\_Omaha\_v01\_001.jpg UNO0012\_Boys\_of\_Omaha\_v01\_001.jpg

...

UNO0012 Boys of Omaha v01.pdf

http://digitalpowrr.niu.edu/

### **Analyzing/OCR**



http://www.cvisiontech.com/library/ocr/fast-ocr/fast-pdf-ocr.html

#### What is PDF/A?

**PDF/A-1a:** The PDF/A-1a requires tagging for structure as well as Unicode character maps for fonts. The objective for PDF/A-1a includes the goals for PDF/A-1b and accessibility for



physically impaired users. The tags for accessibility enable screen readers to provide some form of description for images.

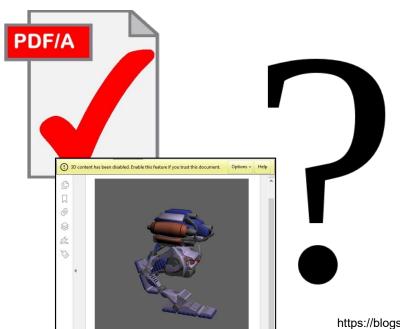
**PDF/A-1b**: The PDF/A-1b variant is for content that has no accessibility tagging; it's useful for scanned documents. The objective for PDF/A-1b is to ensure reliable reproduction of the visual appearance of the document.

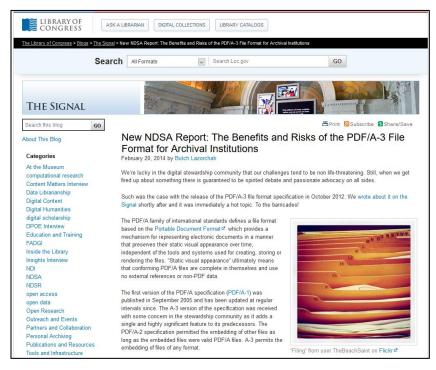
**PDF/A-2**: The PDF/A-2 is the second part of the standard and addresses some of the new features added with versions 1.5, 1.6, and 1.7 of the PDF Reference.

**PDF/A-3**: The PDF/A-3 allows for the embedding of arbitrary file formats such as XML, CSV, and CAD into PDF/A as complete archived objects.

### **Digitization Workflow**

### - (3) Convert





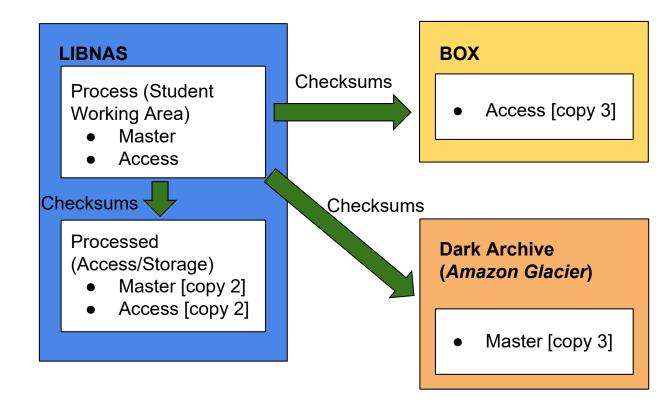
https://blogs.loc.gov/thesignal/2014/02/new-ndsa-report-the-benefits-and-risks-of-the-pdfa-3-file-format-for-archival-institutions/

#### 3-2-1 Rule

3: Make 3 copies

**2:** At least 2 of the copies should be on 2 different types of storage media.

**1:** Store 1 of the copies in a *different location* from the other 2 copies.



#### **Goals of Long-term Digital Preservation**

- Availability
- Identity
- Persistence
- Renderability
- Understandability
- Authenticity



#### **Preservation Metadata**

**Provenance:** Preservation metadata should record information bearing on the custodial history of the digital object, potentially stretching back to the time of the object's creation, and moving forward through successive changes in physical custody and/or ownership.

**Authenticity**: Preservation metadata should include information sufficient to validate that the archived digital object is in fact what it purports to be, and has not been altered, either intentionally or unintentionally, in an undocumented way.

**Preservation activity**: Preservation metadata should document the actions taken over time to preserve the digital object, and record any consequences of these actions that impact the look, feel, or functionality of the object.

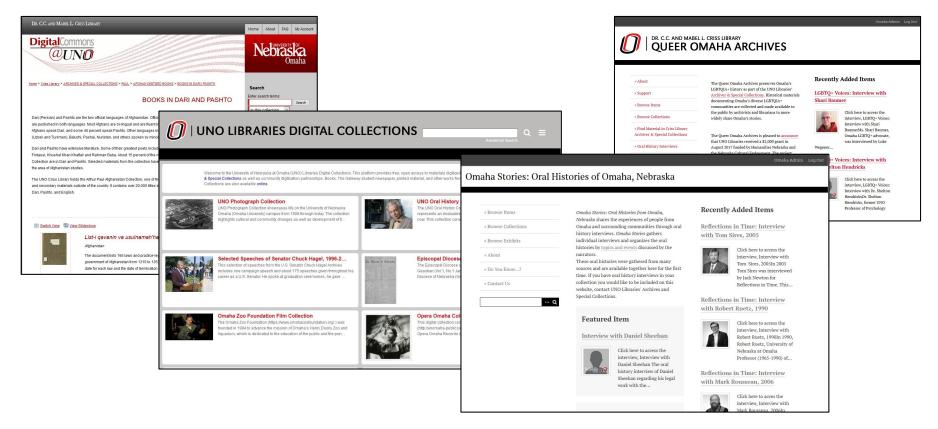
**Technical environment**: Preservation metadata should describe the technical requirements, such as hardware, operating system, and software applications, needed to render and use the digital object in the state in which it is currently stored in the repository.

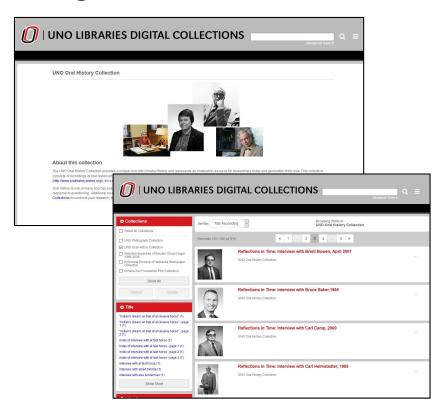
**Rights management**: Preservation metadata should record any binding intellectual property rights that limit the repository's powers to take action to preserve the digital object, and to disseminate the object to current and future users.

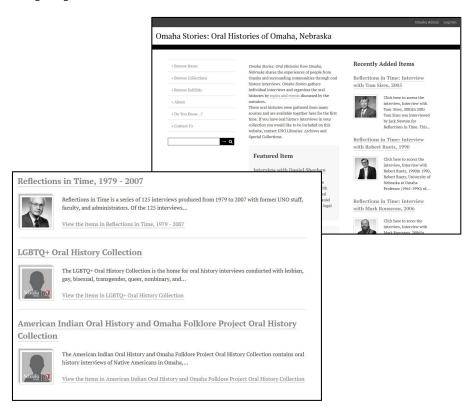
#### Levels of Digital Preservation https://ndsa.org/activities/levels-of-digital-preservation/

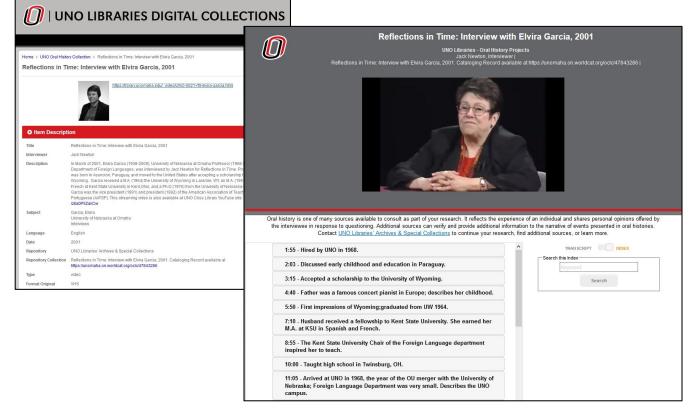
	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)	
Storage and Geographic Location	Two complete copies that are not collocated     For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	Al least three complete copies     At least one copy in a different geographic location     Document your storage system(s) and storage media and what you need to use them  - Check fixity on all	At least one copy in a geographic location with a different disaster threat     Obsolescence monitoring process for your storage system(s) and media	At least three copies in geographic locations with different disaster threats     Have a comprehensive plan in place that will keep flies and metadata on currently accessible media or systems	
File Fixity and Data Integrity			- Check fixity of content at fixed intervals intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	Check fixity of all content in response to specific events or activities     Ability to replace/repair corrupted data     Ensure no one person has write access to all copies	
Information Security - Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files		- Document access restrictions for content	Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs	
Metadata - Inventory of content and its storage location - Ensure backup and non-collocation of inventory		Store administrative metadata     Store transformative metadata and log events	netadata technical and descriptive metadata netadata and log		
File Formats  - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs		- Inventory of file formats in use	- Monitor file format obsolescence issues	Perform format migrations, emulation and similar activities as needed	

- Storage and Geographic Location: between Level 1 and Level2
- File Fixity and Data Integrity:
   Level 1
- Information Security: Level 1
- Metadata: Level 1
- File Formats: Level 2





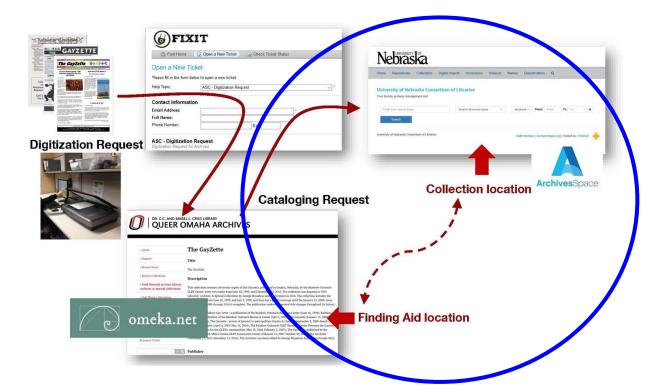






# Oral History Metadata Synchronizer

http://www.oralhistoryonline.org/



### **Questions**





# Thank you!

Corinne Jacox,
Catalog/Reference Librarian,
Creighton University Law Library
corinnejacox@creighton.edu
Creighton
UNIVERSITY

School of Law

Yumi Ohira,
Digital Initiatives Librarian,
University of Nebraska at Omaha Libraries
yohira@unomaha.edu



