The PREMIS of our not so SIMP-le story

Implementing preservation metadata using homegrown and vendor solutions

Jeremy Myntti, Head of Digital Library Services Tawnya Keller, Interim Assistant Head of Digital Preservation



Digital Library at the University of Utah

CONTENTdm for 15 years

Migrating to a Hydra-like platform using Fedora and Solr

450+ collections

- 2.5 million digital objects
- 1.8+ million digital newspaper pages



Need to implement preservation system

- 1st major digitization project: 2000
- Digital Preservation Archivist hired: 2008
- It's GROWING! 250TB to be archived now, growing each day

• Audio-visual collections



• Born-digital collections



Decision to implement Rosetta

Evaluated 4 different systems



_ __ _

@rchivematica.





Our need for developing an ingest system to work with both CDM and Rosetta

Rosetta purchase

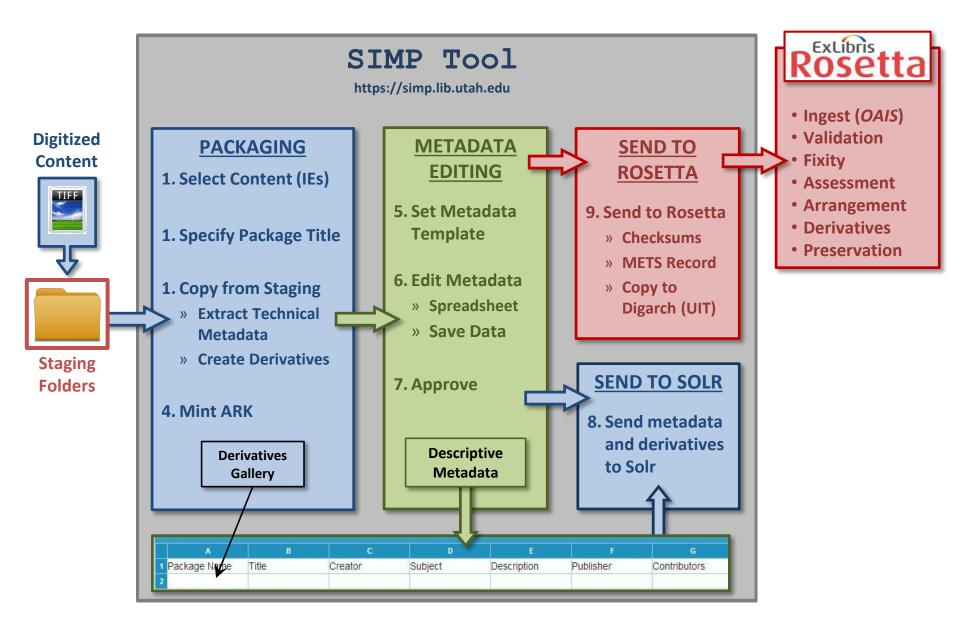
_ _ _

• Disconnect between sales reps and developers

The big problem

• SYNCHRONIZATION between Rosetta and DAM

Digital Operations Metadata Services Digital Preservation



SIMP Tool (Submission Information & Metadata Packaging Tool)

Browse	Servers	Assess Packages	Adm	in Queue (0)	Phalcon			Sandbox	Signout u0468989
Invert	Select all	Select with	out ARK	Select without tem	plate Select uned	lited Select unser	nt Select unapproved	Select cancelled	Select none
0 selec	cted								
Rosetta (I	rosetta)	▼ Se	et Metadata	Template Create AR	s Edit Metadata Ap	oprove Download Meta	adata Select Destination	 Send to Roset 	ta Send to Fedora
	Package Name		▲ Titl	e	Metadata Template		Approved By #ARK #	Rosetta Created Status By	Status 🝦
					/uu_uoh	•			
	132200101_000	6me5x	Dor	ald Adams, Uranium Or	al /uu_uoh	u0628012	ark:/87278/s6tq87kr	u0514441	G by u0628012
	132200102_010	6me5x	Rot	ert Anderson, Uranium	D /uu_uoh	u0628012	ark:/87278/s6pz7gwg	u0514441	P by u0628012
	132200103_020	6me5x	Jeri	y Anderson, Uranium O	a /uu_uoh	u0628012	ark:/87278/s6k95fjb	u0514441	P by u0628012
	132200104_030	6me5x	Jim	Anderson, Uranium Ora	I /uu_uoh	u0628012	ark:/87278/s6fj4pv2	u0514441	🔒 by u0628012
	132200105_040	6me5x	Pea	rl Baker, Uranium Oral H	li /uu_uoh	u0628012	ark:/87278/s69s3z2v	u0514441	🔒 by u0628012
	132200106_050	6me5x	Rot	ert Baldwin, Uranium Or	a /uu_uoh	u0628012	ark:/87278/s661369c	u0514441	🔒 by u0628012
	132200107_060	6me5x	Har	old Barton, Uranium Ora	I /uu_uoh	u0628012	ark:/87278/s6282fj6	u0514441	🔒 by u0628012
	132200108_070	6me5x	Ken	neth Beach, Uranium Oi	a /uu_uoh	u0628012	ark:/87278/s6xh1z58	u0514441	🔒 by u0628012
	132200109_080	6me5x	Bill	Joe Begay, Uranium Ora	I /uu_uoh	u0628012	ark:/87278/s6sr16gc	u0514441	🔒 by u0628012
	132200110_090	6me5x	Mar	ion E. Benedict, Uraniun	n /uu_uoh	u0628012	ark:/87278/s6p2954p	u0514441	🔒 by u0628012
	132200111_100	6me5x	Mic	hael Benson, Uranium O	r /uu_uoh	u0628012	ark:/87278/s6j98dfd	u0514441	🔒 by u0628012
	132200112_110	6me5x	132	2_001_12	/uu_uoh	u0628012	ark:/87278/s6dj7np7	u0514441	P by u0628012
	132200113_120	6me5x	132	2_001_13	/uu_uoh	u0628012	ark:/87278/s68s6wwm	u0514441	P by u0628012
	132200114_130	6me5x	132	2_001_14	/uu_uoh	u0628012	ark:/87278/s651656f	u0514441	P by u0628012
	132200201_000	6mfk4	132	2_002_01	/uu_uoh	u0628012	ark:/87278/s6185djm	u0514441	P by u0628012

Drawes Canvers Access Daskages Admin Quays (A) Dhalaan Candhay Cignayt y0/	C0000
Browse Servers Assess Packages Admin Queue (0) Phalcon Sandbox Signout u04	00303

/ mnt / LOCKER2 / digops / Univ_of_Utah_-_Alan_K_Engen_Papers / 1601_21_01

Invert	Name	Size	Туре
	1601_21_01_001.tif	197.2M	image/tiff
	1601_21_01_002.tif	197M	image/tiff
	1601_21_01_003.tif	211.4M	image/tiff
	1601_21_01_004.tif	213M	image/tiff
	1601_21_01_005.tif	212M	image/tiff
	1601_21_01_006.tif	211.6M	image/tiff
	1601_21_01_007.tif	211.9M	image/tiff
	1601_21_01_008.tif	208.9M	image/tiff
	1601_21_01_009.tif	211.3M	image/tiff
	1601_21_01_010.tif	211.8M	image/tiff
	1601_21_01_011.tif	213.3M	image/tiff
	1601_21_01_012.tif	209M	image/tiff
	1601_21_01_013.tif	211.5M	image/tiff
	1601_21_01_014.tif	212M	image/tiff

e Se	rvers Assess	s Packages Admin	Queue (0)	Phalcon			Sandbox	Signout u046898	9
earch		Replace	Rep	blace All Rep	place Selected	Case se	nsitive 🗆 Regular expressions		
	Package Name		Title		Alternative title	Links to Media	Creator	Contributor	F
1	136900102_01o6f19r	Esperanza and Gavino Aguayo,	No. 1, Hispanic Oral Histo	ories, Acon 1369			Aguayo, Esperanza, 1932-; Aguayo, Gavino;	Kelen, Leslie G., 1949-;	
2	136900103_02o6f19r	Esperanza and Gavino Aguayo,	No. 2, Hispanic Oral Histo	ories, Accn 1369			Aguayo, Esperanza, 1932-; Aguayo, Gavino;	Kelen, Leslie G., 1949-;	
3	136900104_03o6f19r	Rebecca Florez Alvera, Hispani	: Oral Histories, Accn 136	9			Alvera, Rebecca F., 1925-2008;	Kelen, Leslie G., 1949-;	
4	136900105_04o6f19r	Robert Archuleta, Hispanic Oral	Histories, Accn 1369				Archuleta, Robert, 1930-;	Kelen, Leslie G., 1949-;	

-

F II

Save Save and Unlock Autosave @

Export Metadata For CDM

Import Metadata

€ 🔲

Successfully saved 0 row(s) at 10:42:14 AM

SIMP Tool Processes

Lennox and Catherine Tierney Photo:
Set Metadata Template
Create ARKs

Edit Metadata Approve Download Metadata

Select Destination	۳	Send
Select Destination		
Send to "DHA_Collections"		
Send to "JWML_AV"		
Send to "JWML_CollDev"		
Send to "JWML_IR"		
Send to "JWML_Image"		
Send to "Legacy"		
Send to "SPC_AV"		:
Send to "SPC_Manuscripts"		
Send to "SPC_Photographs"		
Send to "SPC_RareBooks"		
Send to "UDN"		
Send to "USHS_AV"		
Send to "USHS_Collections"		
Send to "USHS_SiteForms"		

Send to Fedora Send to Solr Delete

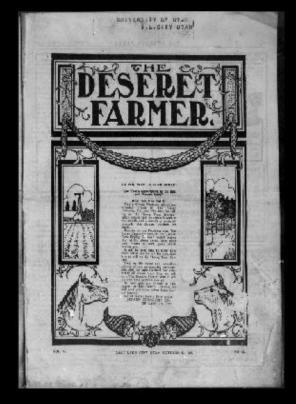
Rosetta view

10	Dete
1704	Depend Farmer, 1908-07-11
c identifier	anc/8727856Mbts
escription	
fornat Identifiar Type	+SPO, internal identifier Value+23825
demail Identifier Type	-PiD, Internal Identifier Value-FL252980
demail identifier Type	-Depositiet0, Wernal Identifier Value-24306
nty Type=SHA1, Bully	Value=740e75956881453225928709155d95a3530e
eth Type+MD5, fixtu'	raive=258+5936+75856+54988+8589442593
ity Type+CRC32, to	6 Value-a 10 dB/d
umat Registry John	6353. format Versionix
ype-image byteOrde	r, Value-Mite-endian
paking creation	Date: Value+2010/06/03 00:22:25
ype-image docume	etitiame, Value-00237295384
	ateValue, Value+(255)
yes-image mindam	cieValue, Value+(0)
ype-mage newSub	FGeT/pe, Valuer(single page of multi-page image)
ype-mischriege titsl	PerSample, Yalve-(3)
penniscimage byte	Order, Value Hittle-endlah
vpentracimage clas	a. Value+class edu harvard hul dis jhove feixolmageMetada
ypennisolmage colo	orflipace. Value+black is zero
ge-nipolmage.com	pressionScheme, Valveruncompressed
vpennisolmage date	TrmeCreated, Value=2010-06-03T00 22:25
pa-nizoimage ima	peLength, Value-5883



DAM view

Title	Deseret Farmer, 1908-10-10
Type	newspaper
Date	1908-10-10
Paper	Deseret Farmer
Rights	Material in the public domain. No restrictions on use.
Publisher	Digitized by: University of Utah
ARK	ark:/87278/s6mg8jxp



Initial decision for minimal descriptive metadata in Rosetta

_ _ _

Tag	Data
Title	39222002396120
Identifier	ark:/99999/fk4sq90d04
Relation	/testkm

Changed that to include MWDL fields

_ __ _

Tag	Data
Title	39222004416967
Identifier	ark:/87278/s6zk7gk4
Relation	/UT-ArchSite
Publisher	
Creator	W. R. Latady
Subject	
Description	
Is Part Of	

Preservation metadata currently created in Rosetta

Information such as

- objectIdentifier
- objectCategory
- objectCharacteristics
- format

_ _ _

- storage
- eventIdentifier
- eventType

NDSA guidelines: where we are now and where we are headed

	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	 At 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 	- 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 files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	Check fixity on all ingests Use write-blockers when working with original media Virus-check high risk content	Check fixity of content at fixed 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	- Store standard technical and descriptive metadata	- Store standard preservation metadata
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

NDSA metadata guidelines

- Level 1 (Protect your data)
 - Inventory of content and its storage location
 - Ensure backup and non-collocation of inventory
- Level 2 (Know your data)
 - Store administrative metadata
 - Store transformative metadata and log events
- Level 3 (Monitor your data) -- We are here right now
 - Store standard technical and descriptive metadata
- Level 4 (Repair your data) -- Working towards this level
 - Store standard preservation metadata

Preservation metadata for the future in Rosetta

NDSA levels ---> work to become PREMIS conformant

Information related to

- environment
- Event

_ _ _

- linking
- soMuchMore
- itsOverwhelmingHelp

Questions?

_ _ _

Jeremy Myntti, Head of Digital Library Services jeremy.myntti@utah.edu

Tawnya Keller, Interim Assistant Head of Digital Preservation <u>tawnya.keller@utah.edu</u>

