

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



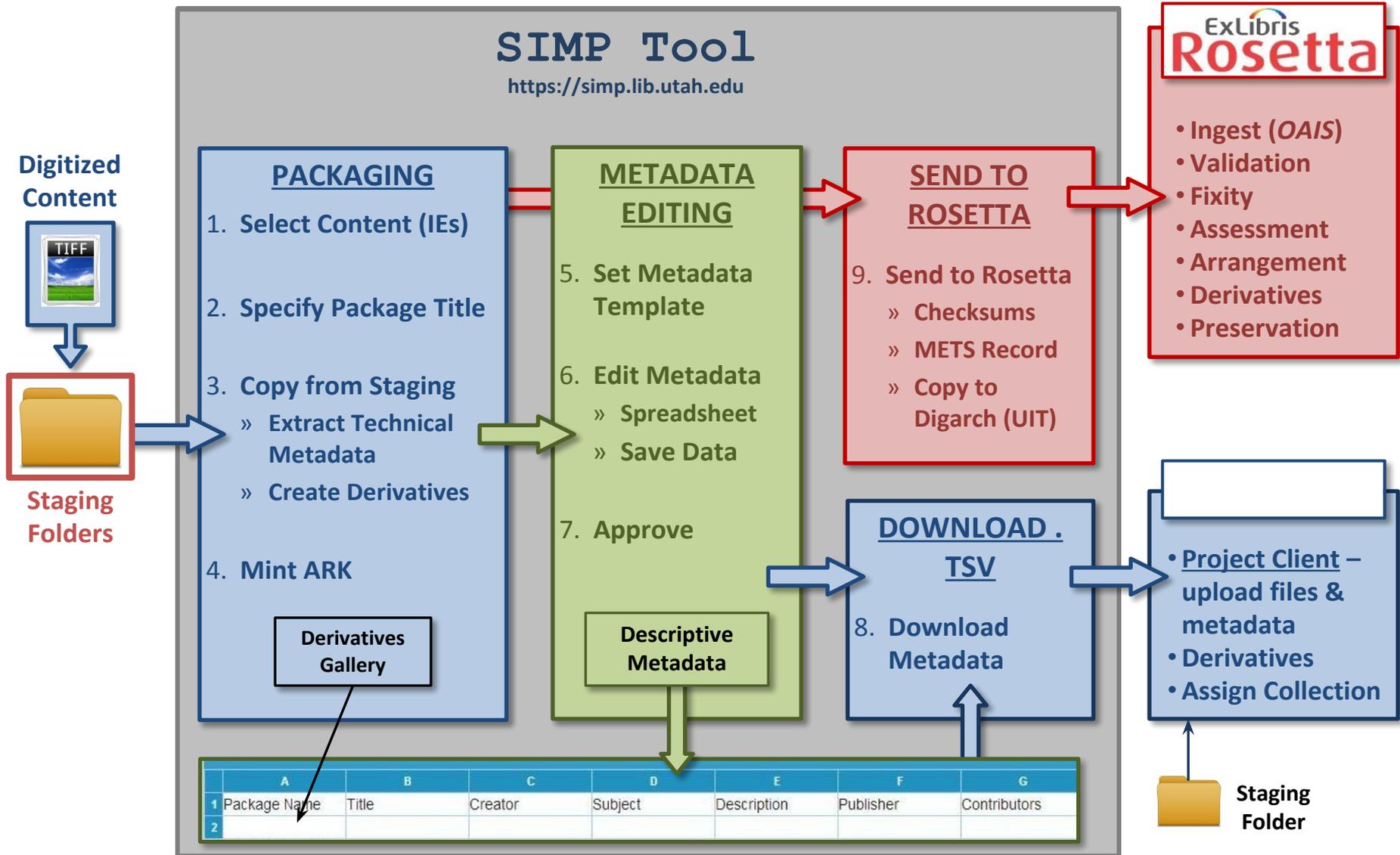
Our need for developing an ingest system to work with both CDM and Rosetta

Rosetta purchase

- Disconnect between sales reps and developers

The big problem

- SYNCHRONIZATION between Rosetta and DAM



SIMP Tool

(Submission Information & Metadata Packaging Tool)

Browse Servers		Assess Packages		Admin		Queue (0)		Phalcon		Sandbox		Signout u0468989					
Invert	Select all	Select without ARK	Select without template	Select unedited	Select unsent	Select unapproved	Select cancelled	Select none									
0 selected																	
Rosetta (rosetta)		Set Metadata Template		Create ARKs		Edit Metadata		Approve		Download Metadata		Select Destination		Send to Rosetta		Send to Fedora	
Package Name	Title	Metadata Template	Last Editor	Approved By	ARK	Rosetta Status	Created By	Status									
<input type="checkbox"/>		/uu_uoh															
<input type="checkbox"/>	132200101_00o6me5x	Donald Adams, Uranium Oral...	/uu_uoh	u0628012	ark:/87278/s6tq87kr	u0514441	by u0628012										
<input type="checkbox"/>	132200102_01o6me5x	Robert Anderson, Uranium O...	/uu_uoh	u0628012	ark:/87278/s6pz7gwg	u0514441	by u0628012										
<input type="checkbox"/>	132200103_02o6me5x	Jerry Anderson, Uranium Ora...	/uu_uoh	u0628012	ark:/87278/s6k95fjb	u0514441	by u0628012										
<input type="checkbox"/>	132200104_03o6me5x	Jim Anderson, Uranium Oral ...	/uu_uoh	u0628012	ark:/87278/s6fj4pv2	u0514441	by u0628012										
<input type="checkbox"/>	132200105_04o6me5x	Pearl Baker, Uranium Oral Hi...	/uu_uoh	u0628012	ark:/87278/s69s3z2v	u0514441	by u0628012										
<input type="checkbox"/>	132200106_05o6me5x	Robert Baldwin, Uranium Ora...	/uu_uoh	u0628012	ark:/87278/s661369c	u0514441	by u0628012										
<input type="checkbox"/>	132200107_06o6me5x	Harold Barton, Uranium Oral ...	/uu_uoh	u0628012	ark:/87278/s6282fj6	u0514441	by u0628012										
<input type="checkbox"/>	132200108_07o6me5x	Kenneth Beach, Uranium Ora...	/uu_uoh	u0628012	ark:/87278/s6xh1z58	u0514441	by u0628012										
<input type="checkbox"/>	132200109_08o6me5x	Bill Joe Begay, Uranium Oral ...	/uu_uoh	u0628012	ark:/87278/s6sr16gc	u0514441	by u0628012										
<input type="checkbox"/>	132200110_09o6me5x	Marion E. Benedict, Uranium ...	/uu_uoh	u0628012	ark:/87278/s6p2954p	u0514441	by u0628012										
<input type="checkbox"/>	132200111_10o6me5x	Michael Benson, Uranium Or...	/uu_uoh	u0628012	ark:/87278/s6j98dfd	u0514441	by u0628012										
<input type="checkbox"/>	132200112_11o6me5x	1322_001_12	/uu_uoh	u0628012	ark:/87278/s6dj7np7	u0514441	by u0628012										
<input type="checkbox"/>	132200113_12o6me5x	1322_001_13	/uu_uoh	u0628012	ark:/87278/s68s6wmm	u0514441	by u0628012										
<input type="checkbox"/>	132200114_13o6me5x	1322_001_14	/uu_uoh	u0628012	ark:/87278/s651656f	u0514441	by u0628012										
<input type="checkbox"/>	132200201_00o6mfk4	1322_002_01	/uu_uoh	u0628012	ark:/87278/s6185djm	u0514441	by u0628012										

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

Invert	Name	Size	Type
	--		
	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

Search

Replace

Replace All

Replace Selected

 Case sensitive Regular expressions

	Package Name	Title	Alternative title	Links to Media	Creator	Contributor	Put
1	136900102_01o6f19r	Esperanza and Gavino Aguayo, No. 1, Hispanic Oral Histories, Accn 1369			Aguayo, Esperanza, 1932-; Aguayo, Gavino;	Kelen, Leslie G., 1949-;	
2	136900103_02o6f19r	Esperanza and Gavino Aguayo, No. 2, Hispanic Oral Histories, Accn 1369			Aguayo, Esperanza, 1932-; Aguayo, Gavino;	Kelen, Leslie G., 1949-;	
3	136900104_03o6f19r	Rebecca Florez Alvera, Hispanic Oral Histories, Accn 1369			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-;	

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](#)

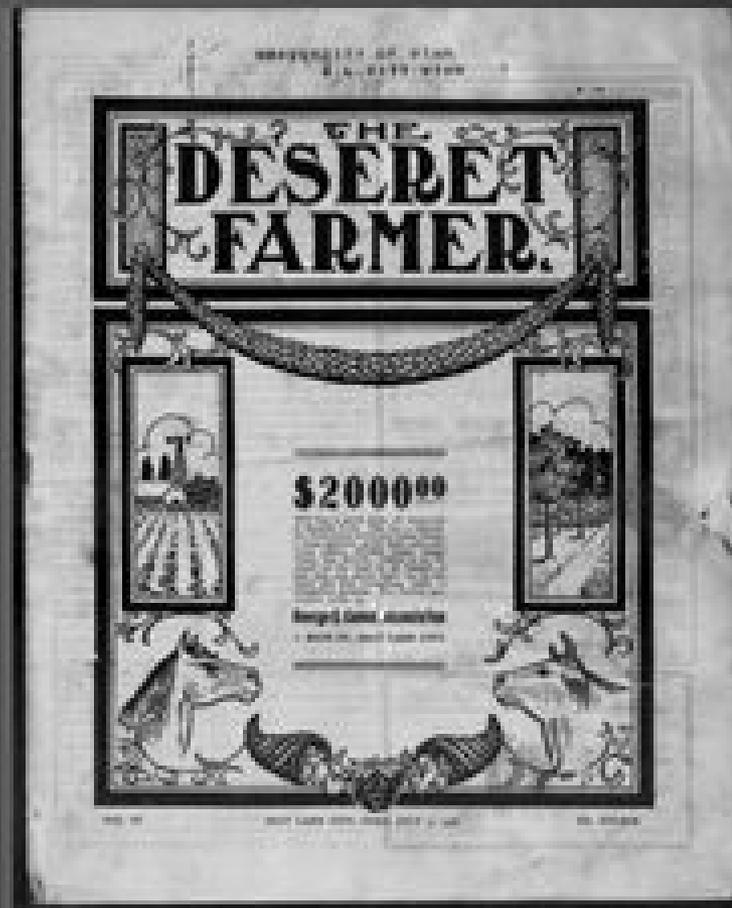
Select Destination ▾ [Send to Rosetta](#)

- 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_Photos"
- 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](#)

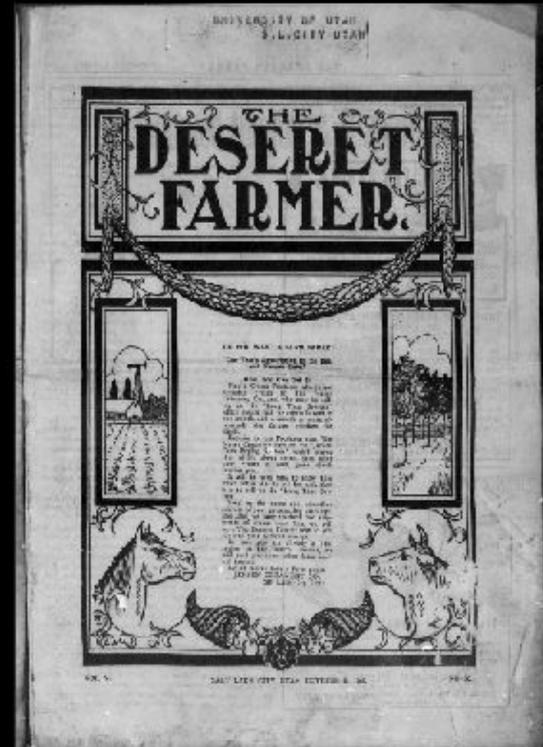
Rosetta view

Tag	Data
code	Deseret Farmer, 1909-07-11
identifier	urn:0727516902
description	
formatRegistryId	urn:mpeg:mpeg211:format:211
image:byteOrder	little-endian
image:creationDate	2010-06-03 00:22:28
image:documentId	00237295384
image:height	250
image:width	0
image:hasSinglePage	single page of multi-page image
image:bitsPerSample	0
image:byteOrder	little-endian
image:colorSpace	edc:harvard-hul:dc:jhove:fiscolmage:Metadata
image:compressionScheme	uncompressed
image:dateTimeCreated	2010-06-03T00:22:28
image:imageLength	5693



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

Table 1: Version 1 of the 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	<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content 	<ul style="list-style-type: none"> - Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content 	<ul style="list-style-type: none"> - 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 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files 	<ul style="list-style-type: none"> - Document access restrictions for content 	<ul style="list-style-type: none"> - Maintain logs of who performed what actions on files, including deletions and preservation actions 	<ul style="list-style-type: none"> - Perform audit of logs
Metadata	<ul style="list-style-type: none"> - Inventory of content and its storage location - Ensure backup and non-collocation of inventory 	<ul style="list-style-type: none"> - Store administrative metadata - Store transformative metadata and log events 	<ul style="list-style-type: none"> - Store standard technical and descriptive metadata 	<ul style="list-style-type: none"> - Store standard preservation metadata
File Formats	<ul style="list-style-type: none"> - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs 	<ul style="list-style-type: none"> - Inventory of file formats in use 	<ul style="list-style-type: none"> - Monitor file format obsolescence issues 	<ul style="list-style-type: none"> - 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
tawnya.keller@utah.edu

