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

---

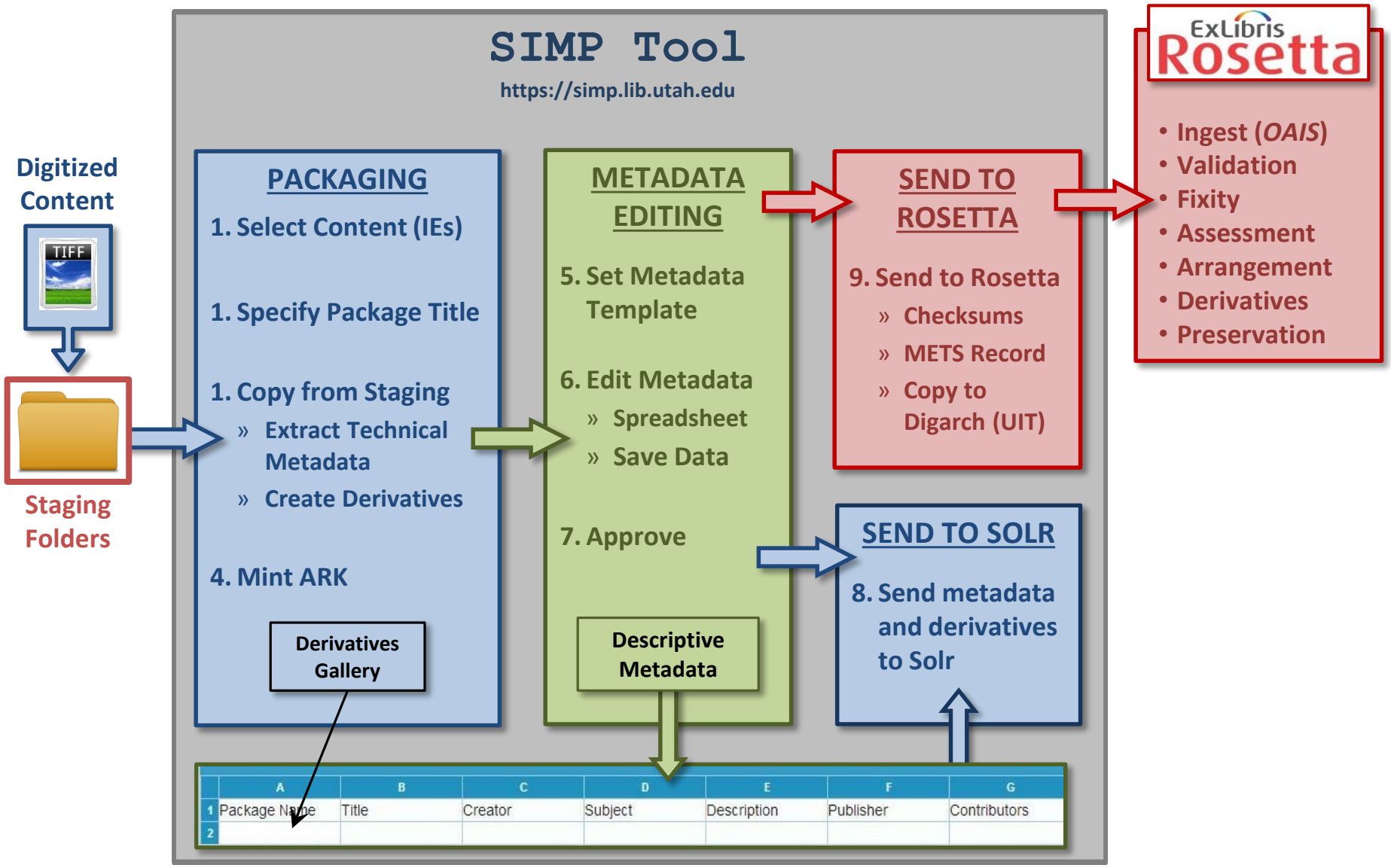
Rosetta purchase

- Disconnect between sales reps and developers

The big problem

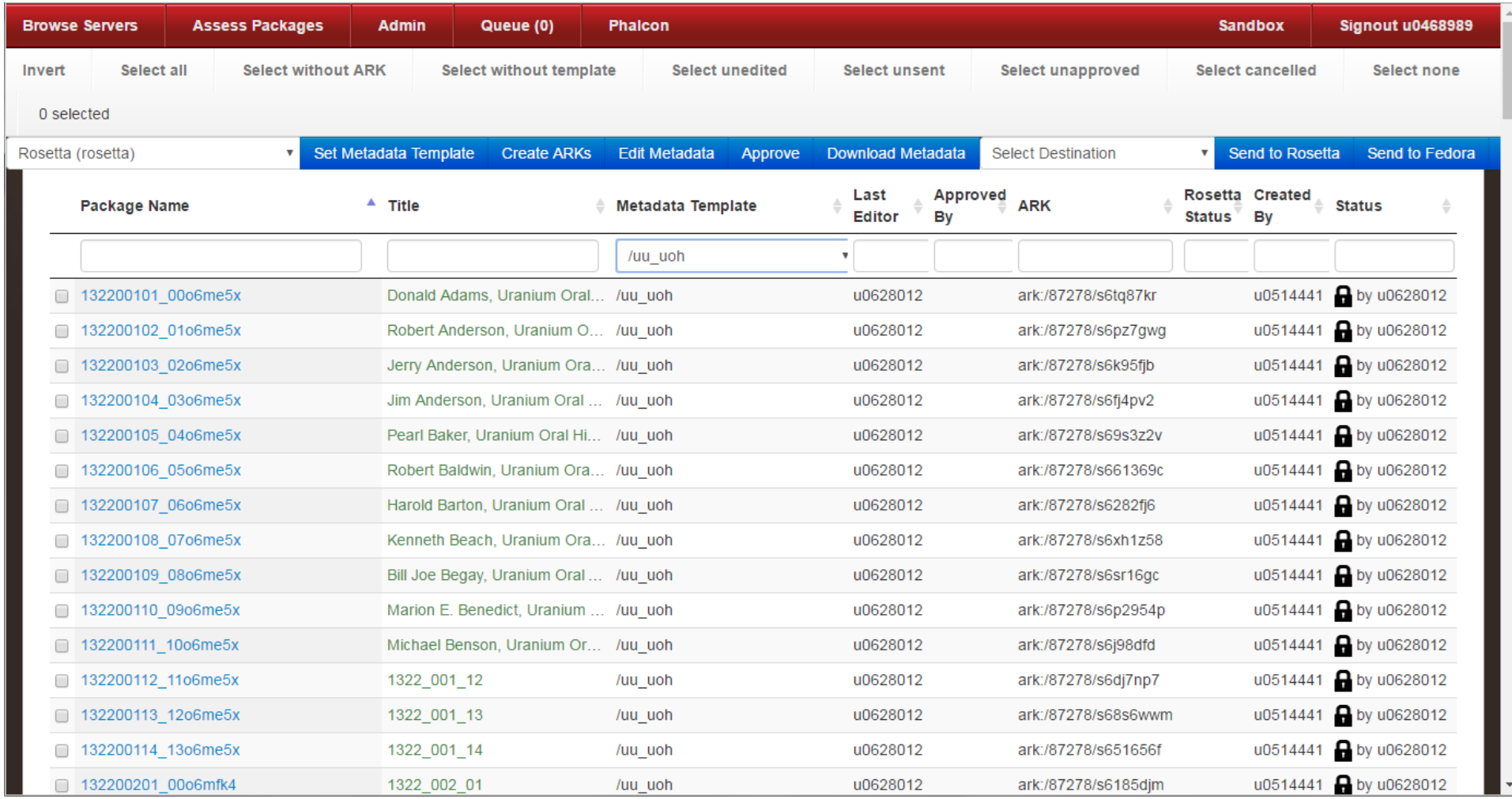
- SYNCHRONIZATION between Rosetta and DAM

Digital Operations
Metadata Services
Digital Preservation



SIMP Tool (Submission Information & Metadata Packaging Tool)

– – –



/ 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	Pub
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-;	

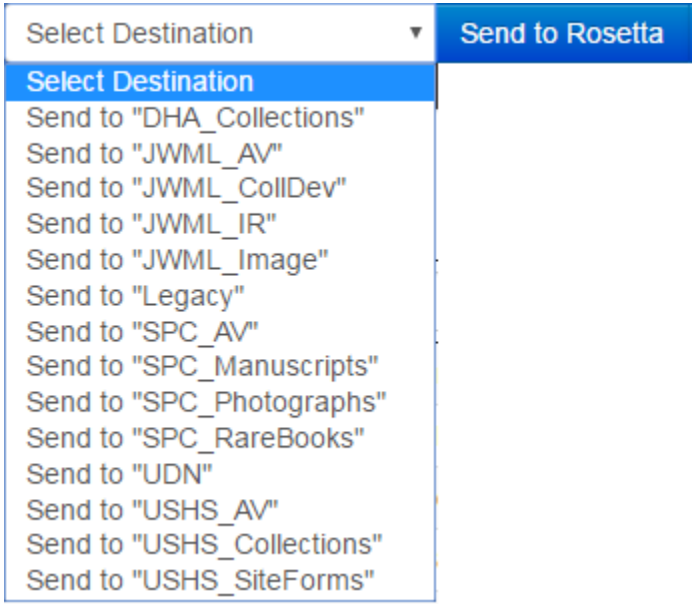
Save | Save and Unlock | Autosave

Export Metadata | For CDM

Import Metadata

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

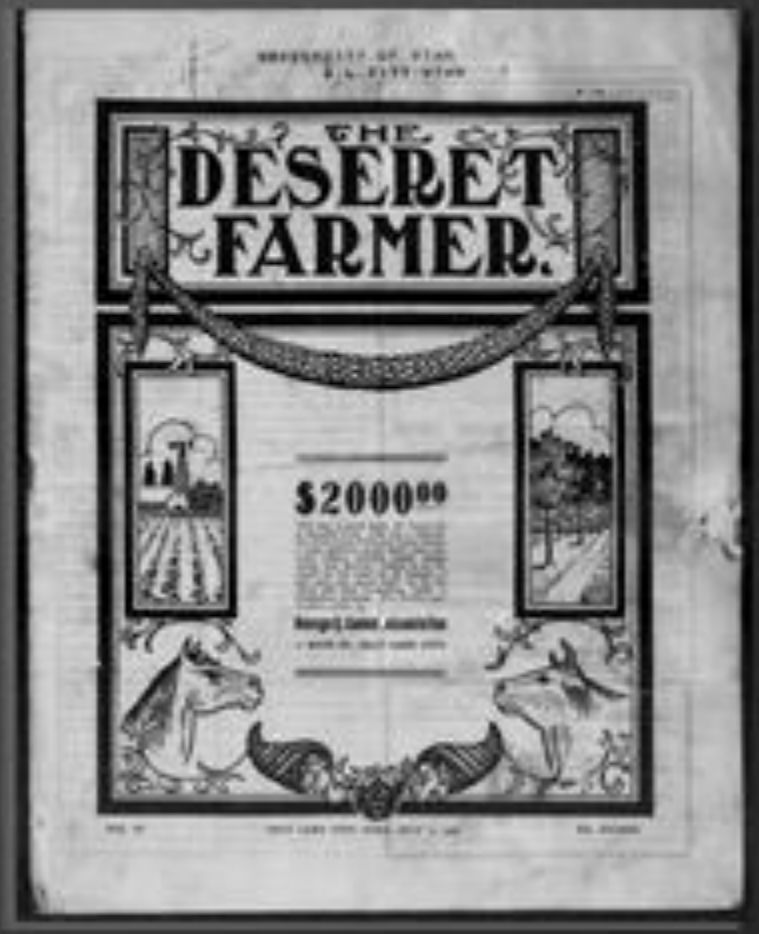
SIMP Tool Processes



Rosetta view

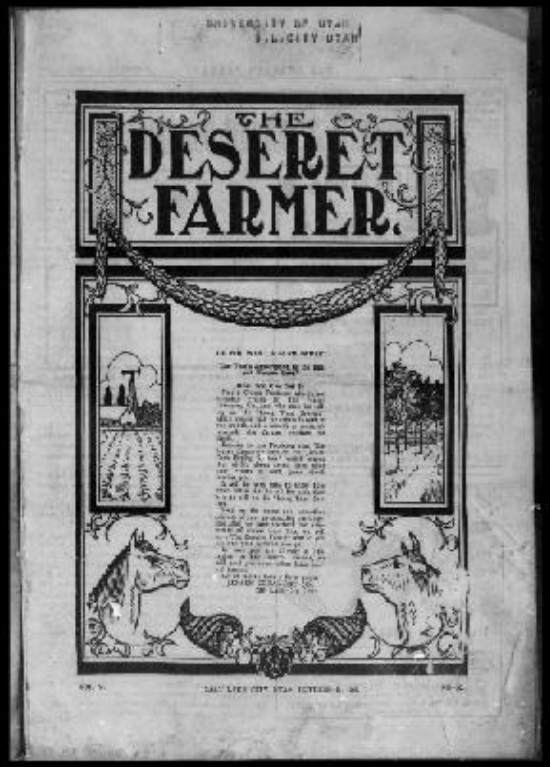
ag	Data
title	Deseret Farmer, 1908-07-11
identifier	ark:/87278/s9hf0zzs
escriptions	

internal Identifier Type=SIPID, internal Identifier Value=23825
internal Identifier Type=PID, internal Identifier Value=FL252980
internal Identifier Type=DepositSetID, internal Identifier Value=24306
sity Type=SHA1, fixity Value=74fde75f9c0981453225f0297091c5df99a3530e
sity Type=MD5, fixity Value=256e55de75f55e5498fafa8ffaf425fd
sity Type=CRC32, fixity Value=e10d8fc9
rmat Registry id=fmt/353, format Version=
ype=image.byteOrder, Value=little-endian
ype=image.creationDate, Value=2010-06-03 00:22:26
ype=image.documentName, Value=00237295384
ype=image.maxSampleValue, Value=[255]
ype=image.minSampleValue, Value=[0]
ype=image.newSubFileType, Value=[single page of multi-page image]
ype=nisoImage.bitsPerSample, Value=[8]
ype=nisoImage.byteOrder, Value=little-endian
ype=nisoImage.class, Value=class edu.harvard.hul.ois.jhove.fitsconverter.fitsdat
ype=nisoImage.colourSpace, Value=black is zero
ype=nisoImage.compressionScheme, Value=uncompressed
ype=nisoImage.dateTimeCreated, Value=2010-06-03T00:22:26
ype=nisoImage.imageLength, 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	392220023961 20
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	- 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

http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf

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