Digital Preservation at TTU

For the Texas Digital Library DuraCloud Working Group

Introduction

Even though the Texas Tech University Libraries began scanning digital collections in 2005, the Digital Resources Unit at the TTU library was only formed in 2012 when the digital lab (headed by one manager and one night manager), two metadata librarians, and one digital services librarian were brought together under one supervisor, Joy Perrin, Digital Resources Librarian. In October 2013, a third metadata librarian and a digital stewardship librarian were added to the unit.

Together, we handle the digitization, access, and digital preservation of the main library's local digital collections, the ThinkTech repository (including theses and dissertations), and some of the Southwest Collection's digital collections.

Metadata and Collection Packaging

When the unit was started, there existed what was known as the "metadata bottleneck." From 2009-2011, a few projects took up all of the metadata librarians' attention while the digital lab was still producing other projects. A backlog of a few years' worth of work developed from it. We were able to finally break through the bottleneck in the fall of 2014 after adding the third metadata librarian, adding a digital stewardship librarian to handle the archiving side, employing metadata student assistants (around six at any given time), and refining workflows.

The metadata librarians receive the digital items from the lab in a folder structure laid out in Appendix A. They will add the display copy of each item – usually a PDF – to DSpace and make its metadata record there. From there, they will assemble the collection's manifest, an Excel spreadsheet that connects each item and its description to its repository URI, also described in Appendix A. Until this past summer, they would also download the CSV metadata file for each item from DSpace. Now, in the interest of time and to help break through the bottleneck, the digital stewardship librarian downloads one CSV file for the whole collection later on in the process.

The workflow is slightly different for the electronic theses and dissertations (ETDs). These items will include any drafts of the documents included in Vireo as well as a record of any changes made to the embargo status. Here at the library, we allow students to change the embargo status on their thesis or dissertation at any time, and when they contact us to make those changes, we archive a copy of that email for our records.

Once the collections are complete, the metadata librarians will move them to a folder called Ready to Archive.

Digital Preservation

The digital stewardship librarian first sees collections in the Ready to Archive folder. Her job is then to check that everything is named correctly and in their correct folders. After that, she will move collections over to the archive server. Once the collections are moved, she'll update the manifest for the

10/16/2014

collection and will download the metadata for the collection in a CSV file from DSpace, which she will also add to the archive.
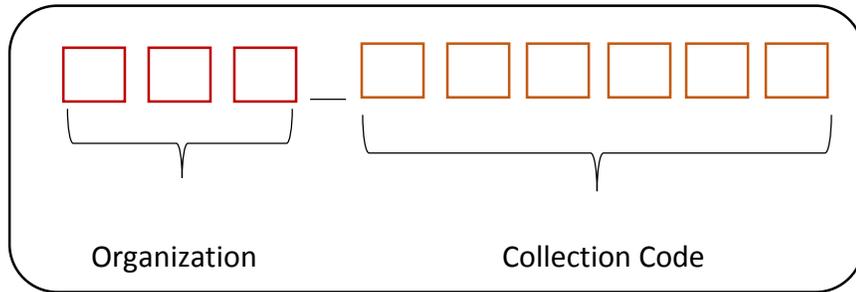
Future for Preservation Metadata

At this point, we are more in the business of preserving metadata than we are of making preservation metadata records. This is not by any means our ideal situation. The digital stewardship librarian's primary goal over the next year, however, is exploring preservation metadata schemas and constructing a workflow that works best for our unit. We are particularly interested in the preservation metadata needed for the ETDs, which would have to include embargo information and records of any changes made.
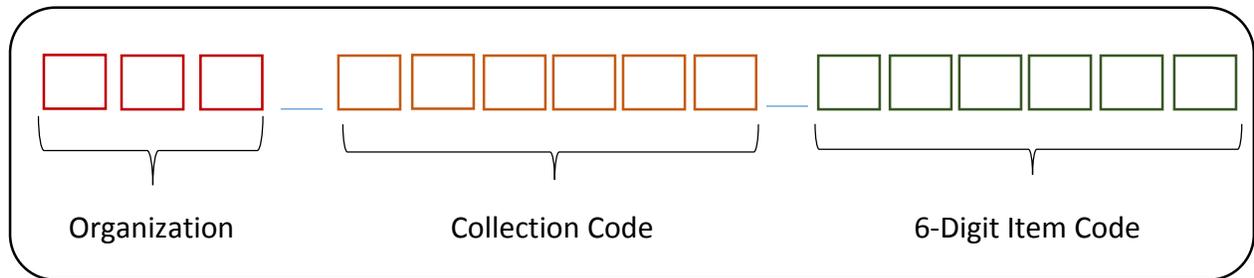
10/16/2014

Appendix A: TTU Dark Archive Workflow

Folder Structure

Main Folder: Collection Title using unique identifier

Organization

Collection Code

Item Folder: each individual item will have its own folder

Organization

Collection Code

6-Digit Item Code

Files

Content Folder

archive

Example:
ttu_swc001_000001_000001.tif

display

Example:
ttu_swc001_000001.pdf

10/16/2014
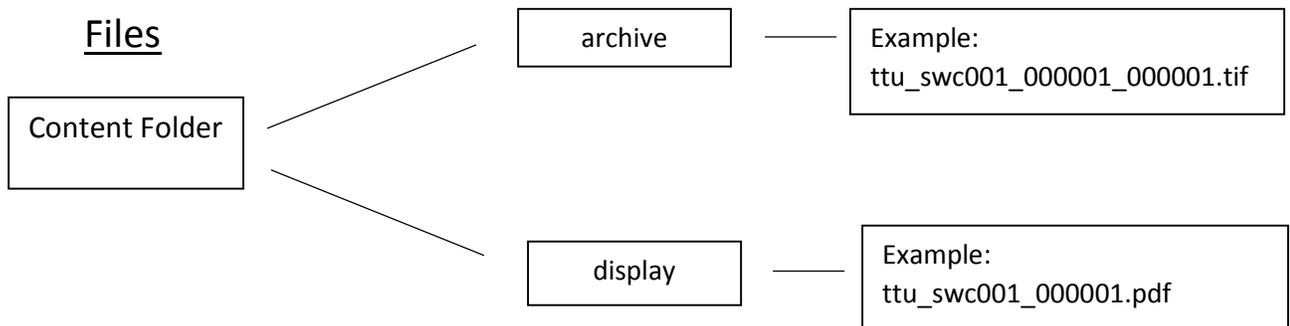
Collection Manifests

Each collection includes an Excel spreadsheet called the manifest that connects each unique item code with its item and URI. Each manifest will differ slightly based on what information is necessary for the records. The ETDs, for example, will include information about embargo statuses, while newspaper collections will include volume and issue numbers. These manifests are vital to the digital preservation process here at TTU. They are saved as 'ttu_[collection code]_manifest.xlsx'.

Example manifest: ETD collection

NEW file NAME	TITLE	LAST NAME	FIRST NAME	Department	DEGREE	DA	EMBAR	NOTES	MD Archiv	URI
ttu_etd001_013451	Effect of ration on growth and development	Allison	James I.	Animal Science	Masters	1976-12		Unrestricted.		http://hdl.handle.net/2346

Collection Metadata

As of fall 2014, we only preserve metadata downloaded at the collection level from DSpace. It is saved in the archive as a CSV file. In each folder, it saved as 'ttu_[collection code]_metadata.csv'.

Example metadata: Hereford Brand collection (from the Southwest Collection DSpace)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	id	collection	dc.contrib	dc.contrib	dc.covera	dc.date.a	dc.date.ac	dc.date.a	dc.date.is	dc.date.is	dc.descrip	dc.descrip	dc.donor	dc.extent	dc.format	dc.format	dc.format	dc.identif	dc.identif	dc.identif	dc.langua
2	56779	10605/503	Landis, De Witt		Deaf Smith County (	2014-03-27T21:00:22	2014-03-27T21:00:22		4/5/1934	Submitte	Weekly newspaper.	10 pages		10 pages	application/pdf			http://hdl.handle.ne	en_US

10/16/2014