


Providing Broader Access: Texas State Retrospective ETD Digitization Project

Todd Peters, Jason Long, Erin
Mazzei, Laura Waugh

Texas Conference on Digital
Libraries

May 2021

Background

- Approximately 6,000 print theses held as second preservation copies in University Archives (about 200 linear feet) which can be disbound and scanned
 - Institutional repository established in 2004 with BePress and migrated to DSpace in in 2012
 - Contracted with 3rd party vendor Atmire to install and manage DSpace Repository
 - Implementation of Vireo for born digital ETD workflows
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.



Legal Issues

- Prior to the Digital Collections Repository, **no license agreements or permissions** were formalized for electronic access
- Print copies available to **view at the Library**
- Authors **retain copyright** and ownership
- Review by Copyright Officer, Texas State University, and Texas State University System attorneys

Solution

Restrict access to
require log-in with
University NetID



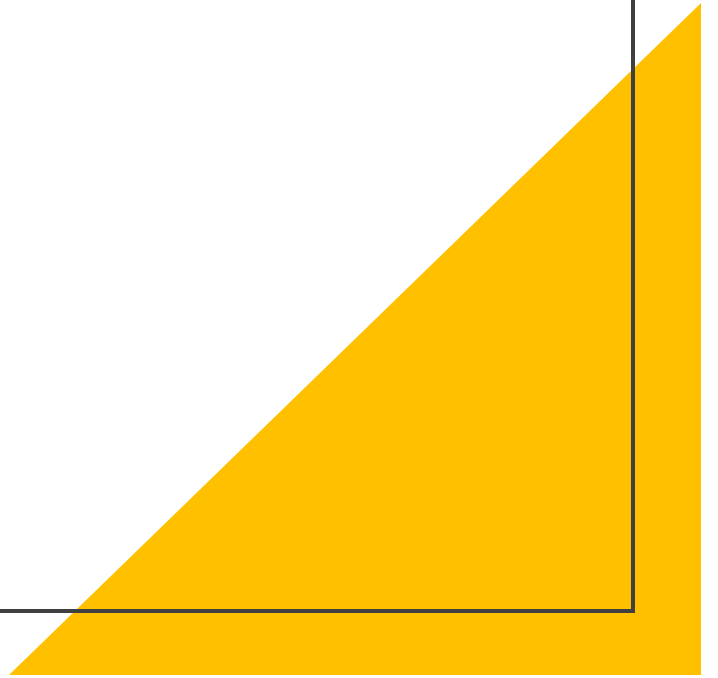
Include a form for
the author to grant
permission for open
online access

Technology Process

- Added new group “Affiliates”
 - Uses Shibboleth attributes
 - Automatically assigns university faculty, staff, and students in “Affiliates”
 - Requires login (NetID and password)
- All other users assigned to group “Anonymous”
 - Existing group for university campus and external
 - All users can view metadata but must login to view document(s)

Technology Process

- Bitstreams in DSpace
 - Edit groups for individual bitstreams
 - Assign to Groups
 - “Affiliates” = restricted access
 - “Anonymous” = open access



DSpace Bitstream Group Assignments

Bitstream Wakefield_Andrea_2004.pdf (81da4d93-2bb9-4c09-a4e6-9d284ee50ffc)

<input type="checkbox"/>	597948	READ	affiliates [Edit]
--------------------------	--------	------	-----------------------------------



Policies for Bundle TEXT (881ab162-4242-4f9e-bbea-354fc565af79)

<input type="checkbox"/>	598018	READ	Anonymous [Edit]
--------------------------	--------	------	----------------------------------



Bitstream Wakefield_Andrea_2004.pdf.txt (b3e903cf-340a-4f42-82da-4f4f978d99fd)

<input type="checkbox"/>	598020	READ	affiliates [Edit]
--------------------------	--------	------	-----------------------------------

Search



☒ Search Digital Collections

☐ This Collection

BROWSE

All of Digital Collections

Communities & Collections

ACCOUNT

Login

STAT

M

S

M

Purification and partial characterization of pokeweed antiviral protein immunoreactive species

Date

2001-05

Author

 Thomas, Staci S.

Metadata

Show full metadata

No abstract prepared.

Citation

Thomas, S. S. (2001). *Purification and partial characterization of pokeweed antiviral protein immunoreactive species* (Unpublished thesis). Southwest Texas State University, San Marcos, Texas.

Uri

<https://digital.library.txstate.edu/handle/10877/13274>

Download

This item is restricted to the Texas State University community. TXST affiliated users can access the item with their **NetID and password authentication**. Non-affiliated individuals should request a copy through their local library's interlibrary loan service.

If this is your thesis or dissertation, you can make it open access. This will allow all visitors to view the document. To request open access, fill out the form linked below:

<https://www.library.txstate.edu/research/digital-collections/Request-Open-Access.html>

Request Open Access

Use this form to request that your thesis or dissertation, which has been digitized and included in the Texas State University Digital Collections repository, be made openly accessible.

This form is intended for Former Texas State Students, as the copyright holder, to authorize the release of their works for open access. If you are not the author, please request through the Interlibrary Loan department at your home library.

Contact digitalcollections@txstate.edu for more information or if you have questions.

Please provide information on the thesis or dissertation.

Thesis or Dissertation Title *

Form Includes:

- Thesis or dissertation title
- Author
- Are you the author?
 - Yes
 - No
- What is your request?
- Contact information:
 - Name
 - Email
 - Phone (optional)

Digitization Workflow



- Print theses/dissertations are disbound
- Scanned with Sheet Feeder Fujitsu fi-6670
 - High-speed color duplex scanning
- All pages are checked/processed and PDFs created
- MarcEdit download of MARC bibliographic record from Library Catalog Z39.50
- Save in folder on shared drive with PDF

Submission to DSpace

- Script created in PHP using [file marc](#)
- Crosswalk MARC to Dublin Core XML file
- Generate ZIP SIP for DSpace ingest
 - Contains PDF and XML file
 - Automatically assigned to “affiliates” group

Parse Marc Record using PHP file_marc

```
if ($value instanceof File_MARC_Control_Field) {
    $val = $value;
} else {
    // Iterate through the subfields in this data field
    foreach ($value->getSubfields() as $code=>$subdata) {
        $spx = "";
        if($dc['element'] == 'title' && $code == 'b')
            $spx = " ";
        if($dc['element'] == 'title') {
            $title .= $title." ".$subdata->getData();
        }
        if($thesis['element'] == 'degree' && $code == 'b')
            $spx = " ";
        if($dc['element'] == 'subject' && $code == 'z' )
            $spx = "--";
    }
}
```

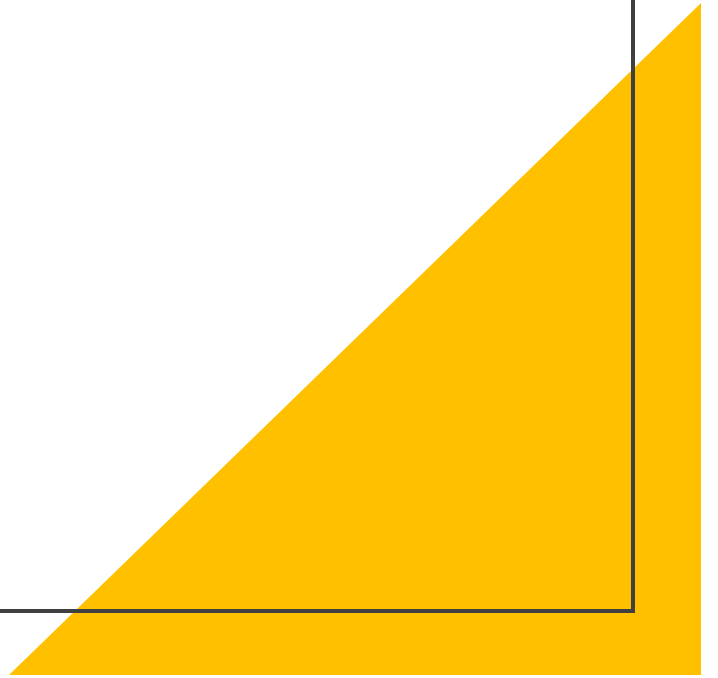
Create SAF format from parsed marc record

<https://wiki.lyrasis.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format>

```
mkdir ($dest_dir);  
copy($source_dir.$pdf_file , $dest_dir . '/' . $pdf_file) or die($dest_dir . '/' . $pdf_file);  
file_put_contents ($dest_dir . '/dublin_core.xml', $dc_xml);  
if($thesis_xml) file_put_contents ($dest_dir . '/metadata_thesis.xml', $thesis_xml);  
if($txstate_xml) file_put_contents ($dest_dir . '/metadata_txstate.xml', $txstate_xml);  
file_put_contents ($dest_dir . '/contents', $pdf_file."\\tbundle:ORIGINAL\\tpermissions:-r  
    affiliates");  
  
rename($source_dir, $export_dir); // Production only  
fputcsv($csv, array($jobNumber,$file,$title, date('m-d-Y h:i:s') , $dept));
```

Submission to DSpace

- Batch Upload Zip files
- 25 items per Zip
 - Manageability
 - Record-keeping
- Post-upload
 - Quality control check
 - Clean-up, as needed
 - Add metadata fields

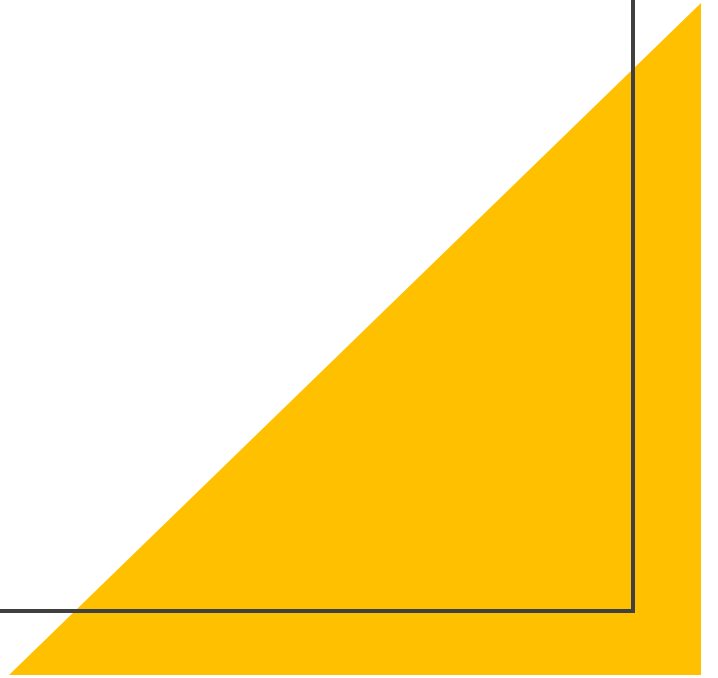


	A	B	C	D
1	Handle	Date Uploaded	Folder (ZIP File)	
61	10877/13258	2021-03-17	1612824193	
62	10877/13259	2021-03-17	1612824193	
63	10877/13260	2021-03-17	1612824193	
64	10877/13262	2021-03-17	1612824193	
65	10877/13263	2021-03-17	1612824193	
66	10877/13264	2021-03-17	1612824193	
67	10877/13265	2021-03-17	1612824193	
68	10877/13266	2021-03-17	1612824193	
69	10877/13267	2021-03-17	1612824193	
70	10877/13272	2021-03-19	1612824199	
71	10877/13273	2021-03-19	1612824199	
72	10877/13274	2021-03-19	1612824199	
73	10877/13275	2021-03-19	1612824199	
74	10877/13276	2021-03-19	1612824199	
75	10877/13277	2021-03-19	1612824199	
76	10877/13278	2021-03-19	1612824199	
< > ≡ 2021 FY 2019-2020 Withdrawn Duplicates Released OA 2019-08 Test Upload +				

Record
Keeping

Metadata

- New field txstate.access with value "restricted"
 - Allows auditing for accidental open access
- Authority control with DSpace “Concepts”
 - Author Concepts
 - Institutional Concepts
- Integration with ORCID
 - Automatic check for potential matches in ORCID
 - Display and link to ORCID, when available





BROWSE

All of Digital Collections

Communities & Collections

ACCOUNT

Login

STATISTICS

Most Popular Items

Statistics by Country

Most Popular Authors

Levy, Taylor M.

Last name:

Levy

First name:

Taylor M.

Institution:

Texas State University

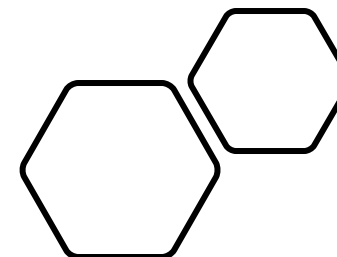
Orcid ID:

 0000-0003-2272-0853

Institution relationship:

Texas State University, Dept. of Health and Human Performance

Show Statistical Information



Lessons Learned

- Testing, as always, is essential
 - Ensure that login link directs users back to the original item in DSpace
 - Batch uploading too many files at once is unmanageable
 - Proper record keeping helps – automated or otherwise

Lessons Learned

- Requests and Fair Use

- Users request access to restricted ETDs for research purposes
- COVID impact? Unclear, but there are frequent requests
- Determination with Copyright Officer:
 - A copy of the PDF can be sent via email
 - Must include statement indicating it is a restricted item
 - Only for personal research purposes

Uploads to Date

1,295	Uploaded	Theses and Dissertations
94	Withdrawn	Creative Writing: Permanent restriction
28	Duplicates	Already existed in repository
4	OA Requests	Author requests to make open access
15	Other Requests	Users seeking access for research

Thank You! Questions?

- Todd Peters, tp09@txstate.edu
- Jason Long, j_l408@txstate.edu
- Erin Mazzei, emazzei@txstate.edu
- Laura Waugh, lwaugh@txstate.edu