Introduction to DSpace

December 8, 2016

Nerissa S. Lindsey, Texas A&M International University
Laura Kane McElfresh, Texas A&M University at Galveston
Intro to DSpace

Today’s training topics include...

- Overview of DSpace and the Texas Digital Library (TDL)
- How to create communities and collections
- How to submit and proceed items through the repository workflow
- How to adjust access options for users and items in DSpace
- Metadata management in DSpace
- Highlights of other DSpace features
Introduction

DSpace and the Texas Digital Library (TDL)
About the Texas Digital Library (TDL)

The Texas Digital Library is a consortium of Texas higher education institutions that builds capacity for preserving, managing, and providing access to unique digital collections of enduring value.

Founded in 2005 by four Association of Research Libraries (ARL) members:

Texas A&M University

Texas Tech University

The University of Houston

The University of Texas at Austin
Infrastructure: large-scale projects, laying the foundation for digital work at member libraries. As a consortium, TDL realizes economies of scale.

DSpace

Texas Data Repository

Open Journals System

...and more

Community: people at member institutions help TDL support its large-scale project work. We develop standards and policies and help inform TDL’s decisions about its large projects.

Working groups

User groups

Training opportunities
DSpace

- Open source software for digital repositories
- Started in 2002 by developers at MIT and HP Labs
- Active development community
- Over 1000 organizations use DSpace
  - Primarily research/higher education

The TDL is a Platinum Member of DuraSpace, the sponsoring organization of DSpace.
DSpace: Features

- Full-text searchable (any text-based file)
- Discovery: search/browse in the DSpace interface, handles (Faceted browse)
- Can handle any type of file; best known for text-based files
  - Files are called “bitstreams”
- Optimized for indexing in Google and Google Scholar
- Persistent URLs (Handle system)
Metadata in DSpace

**Descriptive Metadata**
Flat metadata schema for describing an item
Default: qualified Dublin Core metadata schema; elements & qualifiers pre-configured in source code
However, you can configure multiple schemas/select metadata fields from a mix of configured schemas.

**Administrative Metadata**
Includes preservation metadata; provenance; and authorization policy data
Most administrative metadata is held within DSpace's relational DBMS schema
Some administrative metadata is stored or replicated in Dublin Core records

**Structural Metadata**
Includes relationships between constituent parts of an item; information about how to present the item or its bitstreams to an end user
Example: a thesis in TIFF images, one image per page. Structural metadata tells DSpace 1) that each image is a single page, and 2) the ordering of the TIFF images/pages.
Fairly basic in DSpace: can arrange bitstreams into separate bundles as described above and can designate a primary bitstream
Additional structural metadata can be stored in serialized bitstreams, but DSpace does not currently understand this natively.
Examples of DSpace

- http://repositories.lib.utexas.edu
- http://repository.tamu.edu/
- http://repositories.tdl.org/ttu
- http://repositories.tdl.org/utmb
- https://repositories.tdl.org/uh-ir/
- http://repositories.tdl.org/tamug/
How to Create Communities and Collections
Repository Structure: Communities and Collections

**Community** – highest level of DSpace hierarchy; can contain *sub-communities* and/or *collections*

**Sub-Community** (optional) – if used, contains *collections* or additional nested sub-communities

**Collection** – Contains *items*

**Item** – Contains bitstreams (i.e. files), metadata, and license
Repository structure: Example #1

- Community
  - Collection
    - Item
      - Metadata
      - Bitstreams
      - License
Repository structure: Example #2
The Texas Digital Library uses Shibboleth to manage authentication with TDL services.

Shibboleth lets you log in using your university credentials.

- Training DSpace installation: [https://training-ir.tdl.org/](https://training-ir.tdl.org/)
- There’s no Shibboleth on the training box, so log in with student login credentials:
  - tdl.studentX@gmail.com where 1 ≤ X ≤ 20 (talk amongst yourselves & pick one)
  - password: tdlstudent
Available actions change as you navigate through the interface.
How To:

- Create a Sub-community
- Assign an Administrator to your Sub-community

1. Go to https://training-ir.tdl.org/ (Make sure you are logged in to DSpace.)
2. Navigate to the “UT-RGV Training Community”
3. Click on Create Sub-community
4. Give your Sub-community a name, provide introductory text, and click on Create.
5. Click on the Assign Roles tab
6. Click on Create to assign Administrators
7. Add mcelfrel@tamug.edu &/or nslindsey@tamiu.edu as an Administrator.

Note: Adding an administrative user automatically creates a “Group” called “COMMUNITY_X_ADMIN.”
How To:

- Create a new Collection within your Sub-Community
- Assign an Administrator to your Collection

1. Navigate to the Sub-Community you just created.
2. Click on Create Collection.
3. Give your collection a name, provide some introductory text, and click on Create.
4. On “Assign Roles” tab, click Create next to Administrators.
5. Add anyone you like as an Administrator for the Collection.

Note: Initially, when you add a user as Administrator, the user will appear as “Pending” until you click SAVE.
How To:

- Edit an existing Collection

1. Return to [DSpace Home](#)

2. Navigate to the Collection you just created.


4. Edit any metadata for the collection and upload an image under “Upload new logo.”

5. Click Save updates.
Community Structure in TAMUG’s Repository

Community

Sub-Communities

Collections

Select a community to browse its collections.

- Faculty Scholarly Works
  Scholarly materials created by TAMUG faculty members

- Sargassum Early Advisory System (SEAS)
  Collection of data and reports for advanced forecasting of Sargassum
  (seaweed) landings along the Texas Gulf coast

- Galveston Historic Beach Profiles
  Historic beach profile data for Galveston Island and the north Texas coast, spanning 1966-2008.
  - Beach Profile Monitoring 2002
  - Beach Profile Monitoring 2006
  - Old Profiles 1966-1980
    Historic data from 1966-1980
  - Profile 2007-2008

- Azores historic weather
  Sea level pressure (mb) and composite mean pressure maps by month

- Effects of Sargassum on West End Beaches Galveston 2006
  The effect of sargassum (seaweed) on beach accretion was studied on West Galveston Island

- Liberal Studies Department (LIST)
- Marine Biology Department (MARB)
- Marine Engineering Technology Department (MARR)

Above: Communities and Collections in the Texas A&M at Galveston Repository
Case Study: SEAS Community

Sargassum Early Advisory System Community in the Texas A&M at Galveston Repository

Sargassum Early Advisory System (SEAS)

BROWSE BY

By Issue Date  Authors  Titles  Subjects

Search within this community and its collections:

The beaches of Galveston Island experience an annual beach cast of Sargassum ("seaweed") during late spring/early summer - a season of significant economic input to the island and county. This accumulation inspires much debate among numerous stakeholders including property owners, business owners, recreational fishers, tourists, and local, county, and state governmental entities. The data and reports in this collection study the impact of the sargassum on the beaches including erosion processes, and support forecasting the arrival of large drifts of the sargassum.

Sub-communities within this community

Galveston Historic Beach Profiles
Historic beach profile data for Galveston Island and the north Texas coast, spanning 1966-2008.
Adding Content in DSpace

Submission Workflow
Ingest Process

External SIP → Web Submit UI → In Progress Submission → Item Installer → Archived Item

Batch Item Importer → Workflow (optional)
Starting a new submission

Users with “submit” privileges will see a “Submissions” link under My Account.

Click “Start a New Submission” to begin.

Submissions & Workflow tasks

Submissions

You may start a new submission.

The submission process includes describing the item and uploading the file(s) comprising it. Each community or collection may set its own submission policy.
Submission Steps

Select a Collection
- Only collections on which you have “submit” privileges will appear.

Describe the item (3 screens)
- Title and Date of Publication are required.

Determine access
- Make item private? – Item will not be searchable.
- Set up limited embargo? – Provide future date for access

Upload file(s)
- Upload one or multiple files
- Edit metadata specific to each bitstream, including embargo info.

Complete submission
- Click “Complete submission.”

Agree to license
- Agree to license

If Collection has no workflow steps, and you did not replace any restrictions on access, the item will be immediately available in DSpace.

Review
- Review information and make corrections.
Practice
Submit an Item to your Collection.

1. Click “Submissions”
2. Click “start another submission.”
3. Select a collection and click Next.
4. Proceed through the workflow. Upload one or multiple files from desktop folder.
Editing Items

Moving items to a different collection
Making an item private
Replacing or modifying bitstreams
Reordering bitstreams
Editing item metadata
Reorder Bitstreams

In items with multiple files (i.e. bitstreams), an administrator can reorder the files after submission.

1. Complete submission of item
2. Navigate to collection and item just submitted
3. “Edit this item” => Item Bitstreams tab
4. Use arrows on right side to reorder the files
Editing Item Metadata

1. Navigate to the Item
2. Click “Edit this item” under “Context.”
3. Go to “Item Metadata” tab.
4. Edit existing metadata, or add new fields.

PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data in the correct format. If you are not sure what the format is, please do NOT make changes.

Update
Return
Mapping Items

1. Navigate to the Collection you wish to map items into
2. Item Mapper is under “Context” on the right
3. Search for items and map them in.

*Note: this does not move the items -- it just makes them appear in the “Mapped” collection.*
Practice

Edit and map Items in your Collection

Move Items to a different Collection
Add, change, or delete metadata
Add new bitstreams or delete old ones
Map Items into your Collection
...anything else?
Roles and Workflows

E-People, Groups, Authorization
Roles within DSpace

More privileges

Repository Administrator
Community Administrator
Collection Administrator
Reviewer OR Submitter
Reader

Fewer privileges
E-People and Groups

E-People and Groups are the way DSpace identifies users for the purpose of granting privileges.
E-People

E-Person = User Account

- An E-Person can be granted certain privileges within DSpace.
- In TDL-hosted systems, an E-Person is created when a user logs in for the first time.
Groups

Groups = a list of E-People

- Groups can be granted permissions
- Anyone added to the group gets the permissions granted to the group
- Two default groups in DSpace: Administrator and Anonymous
Roles within DSpace

More privileges

Repository Administrator

Community Administrator

Collection Administrator

Reviewer OR Submitter

Reader

Fewer privileges

Administrator

Community_X_Admin

Collection_X_Admin

Collection_X_Submit

Collection_X_Workflow_Step_1

Collection_X_Workflow_Step_2

Collection_X_Workflow_Step_3

Anonymous (by default)
Managing Groups
(Method 1)

Edit Collection → Assign Roles
- Create a group of Collection Administrators
- Create a group of Submitters
- Create a specified Group who can access materials (default is “anonymous)

To create a group: Click “Create” (or “Restrict), search for and add E-People to the group, click SAVE
Managing Groups
(Method 2)

Access Control → Groups

To create a Group: Click “Click here to add a new Group,” give new Group a name, search for and add E-People to the group, click SAVE.

Note: No privileges are attached to any groups created through this method. But groups created here are available to be authorized in other parts of the interface.
Workflows

Without a Workflow in place, items submitted to a Collection in DSpace will automatically be archived and published.

Workflows allow for one, or multiple, steps for reviewing submissions and editing metadata prior to publication.

- A Workflow can have 1, 2, or 3 steps.
- Each step will have an E-Person Group attached.
Available Workflow Steps

Step 1
- Can accept or reject submission

Step 2
- Edit metadata; accept or reject submission

Step 3
- Edit metadata and publish; cannot reject

Notes: A collection might have one or all of these steps. It could have any one of these steps but not the other two.
A Workflow with all three steps
# Creating a Collection Workflow

## Edit Collection → Assign Roles

- Create a Group for the Workflow step(s) you want.
- A step without a Group does not exist.

<table>
<thead>
<tr>
<th>Step Name</th>
<th>Assigned Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept/Reject Step</td>
<td>none</td>
</tr>
<tr>
<td>Accept/Reject/Edit Metadata Step</td>
<td>none</td>
</tr>
<tr>
<td>Edit Metadata Step</td>
<td>none</td>
</tr>
</tbody>
</table>

**Accept/Reject Step**

The people responsible for this step are able to accept or reject incoming submissions. However, they are not able to edit the submission's metadata.

**Accept/Reject/Edit Metadata Step**

The people responsible for this step are able to edit the metadata of incoming submissions, and then accept or reject them.

**Edit Metadata Step**

The people responsible for this step are able to edit the metadata of incoming submissions, but will not be able to reject them.
Submitter submits item to a Collection with “Step 2” in place. Submitter gets this message:

- An email is sent to every E-Person in the Workflow/Reviewer Group.

Reviewer Group also sees this on their Submissions page:

**Submission complete**

Your submission will now go through the review process for this collection. You will receive e-mail notification as soon as your submission has joined the collection, or if there is a problem with your submission. You may also check on the status of your submission by visiting your submissions page.

Go to the Submissions page

Submit another item

**Submissions & Workflow tasks**

**Unfinished submissions**

These are incomplete item submissions. You may also start another submission.

<table>
<thead>
<tr>
<th>Title</th>
<th>Collection</th>
<th>Submitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untitled</td>
<td>Synthesizers</td>
<td>email: Nerissa Lindsey</td>
</tr>
</tbody>
</table>

Remove selected submissions
Workflow Cont.

Review takes the task and reviews submitted item. Reviewer can edit the item’s metadata, approve or reject the item, or return the item to the pool for another Reviewer to pick up.

Workflow tasks

These tasks are items that are awaiting approval before they are added to the repository. There are two task queues, one for tasks which you have chosen to accept and another for tasks which have not been taken by anyone yet.

Tasks you own

<table>
<thead>
<tr>
<th>Task</th>
<th>Item</th>
<th>Collection</th>
<th>Submitter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>No tasks are assigned to you</td>
</tr>
</tbody>
</table>

Tasks in the pool

<table>
<thead>
<tr>
<th>Task</th>
<th>Item</th>
<th>Collection</th>
<th>Submitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awaiting reviewer’s attention</td>
<td>Korg MS-20</td>
<td>Synthesizers</td>
<td>email: Nerissa Lindsey</td>
</tr>
</tbody>
</table>

Take selected tasks

Submissions being reviewed

These are your completed submissions which are currently being reviewed by collection curators.

<table>
<thead>
<tr>
<th>Title</th>
<th>Collection</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korg MS-20</td>
<td>Synthesizers</td>
<td>Awaiting reviewer's attention</td>
</tr>
</tbody>
</table>
Authorization Policies

VERY specific permissions can be created for e-petons and groups by creating authorization policies at the Collection, Item, or Bitstream Level.

Access Control
- People
- Groups
- Authorizations
## Collection-Level Authorization Policies

<table>
<thead>
<tr>
<th>COLLECTION LEVEL</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD/REMOVE</td>
<td>add or remove items (ADD = permission to submit items)</td>
</tr>
<tr>
<td>DEFAULT_ITEM_READ</td>
<td>inherited as READ by all submitted items</td>
</tr>
<tr>
<td>DEFAULT_BITSTREAM_READ</td>
<td>inherited as READ by Bitstreams of all submitted items. Note: only affects Bitstreams of an item at the time it is initially submitted. If a Bitstream is added later, it does not get the same default read policy.</td>
</tr>
<tr>
<td>COLLECTION_ADMIN</td>
<td>collection admins can edit items in a collection, withdraw items, map other items into this collection.</td>
</tr>
</tbody>
</table>
# Other Authorization Policies

<table>
<thead>
<tr>
<th>Item-Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD/REMOVE</td>
<td>add or remove bundles</td>
</tr>
<tr>
<td>READ</td>
<td>can view item (item metadata is always viewable)</td>
</tr>
<tr>
<td>WRITE</td>
<td>Can modify item</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bundle-Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD/REMOVE</td>
<td>add or remove bitstreams to a bundle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bitstream-Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>READ</td>
<td>view bitstream</td>
</tr>
<tr>
<td>WRITE</td>
<td>modify bitstream</td>
</tr>
</tbody>
</table>
Practice
Assign roles, and create workflows

Create e-People and groups
Assign e-People and groups roles/authorizations
Set up a workflow
...anything else?
Metadata Management in DSpace

• Batch edits & metadata cleanup
• Moving/mapping Items en masse
• Fun with spreadsheets!
How To:

Perform batch operations in your Collection

- Edit metadata
- Move/map Items

Navigate to a Community or Collection

Under “Context”, click “Export Metadata”

Save the .csv file and open it in Excel

Columns MUST have metadata element names as headers

Can take out columns you’re not editing

Do not take out “id” and “collection”

Use a double pipe “||” to include multiple values for an element

Lindsey, Nerissa||McElfresh, Laura
How To:

Perform batch operations (cont.)

Map an item by including multiple values under “collection”

Move an item by changing the “collection”

Save the edited file (make sure it’s .csv)

Under “Content Administration”, click “Import Metadata”

“Choose File”, find your .csv file, and click “Upload CSV”

Approve the changes, and you are done!
Practice

Perform batch operations on repository metadata

(We will walk & talk you through this.)
But wait! There’s more!

Highlights of Other DSpace Features
Statistics in Dspace

Usage statistics can be retrieved for your DSpace instance as a whole, and from individual item, collection and community pages. These Usage Statistics pages show:

- Total page visits (all time)
- Total Visits per Month
- File Downloads (all time)
- Top Country Views (all time)
- Top City Views (all time)

Alternatively, you can connect your repository to Google Analytics which is more robust: https://wiki.duraspace.org/display/DSDOC5x/DSpace+Google+Analytics+Statistics
Harvesting

• DSpace exposes metadata for collection by harvesters using the OAI-PMH protocol.
• DSpace can also harvest metadata and/or objects from other OAI-compliant repositories.
• Harvesting of another collection is configured under “Content Source.”

Documentation: https://wiki.duraspace.org/display/DSDOC5x/OAI
MathJax and other DSpace 5.0 features can be learned about here:
https://tdl-ir.tdl.org/tdl-ir/handle/2249.1/76167
Thank you!

Nerissa Lindsey, Texas A&M International <nslindsey@tamiu.edu>

Laura Kane McElfresh, Texas A&M at Galveston <mcelfrel@tamug.edu>