DSpace 5 Release Webinar

DSpace 5 Working Group
Texas Digital Library
January 20, 2016
About DSpace 5

• DSpace is a product of DuraSpace

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

• No major changes with DSpace version 5, just new features and bug fixes

• DSpace 5 release notes: https://wiki.duraspace.org/display/DSPACE/DSpace+Release+5.0+Status
Timelines and Upgrade Procedure

Upgrades completed so far:
• Texas Tech
• Texas A&M Galveston
• Texas A&M International
• Angelo State University
• Texas Digital Library

Upgrading in January:
• Baylor University
• Sam Houston State
• Texas A&M Corpus Christi
• University of Houston
• UT Arlington
• UT Medical Branch at Galveston
• UT Rio Grande Valley
• UT Southwestern Medical Center
• West Texas A&M
DSpace 5 Working Group

Susan Elkins, *Angelo State University*
Robert Luttrell, *Texas Tech University*
Alexandra Mitchell, *Texas A&M Galveston*
Laura McElfresh, *Texas A&M Galveston*
Nerissa Lindsey, *Texas A&M International*
Le Yang, *Texas Tech University*
Ryan Steans, *Texas Digital Library*
Overview of Updates and New Features

• Audio/Video Embedded Player
• Google Analytics from DSpace Admin UI
• Batch Upload via DSpace UI
• ORCID Lookup Feature
• SHERPA RoMEO
• MathJax
• Additional Features and Bug Fixes
  ▪ Performance Improvements
  ▪ Thumbnail Generation
  ▪ OAI-PMH Harvesting and OpenAIRE
  ▪ Page Not Found Bug Fix
Audio/Video Player

- The TDL programming team embedded HTML5 audio and video elements in DSpace 5.

- The embedded audio/video player is supported on all modern browsers and supports common file types.

- The first file uploaded (or primary file) is displayed at the view of “Simple Item Record.” Additional files are displayed in the “Full Item Record.”
In “Full Item Record” view, files are displayed in-line (next to file).

Description and bitstream information is provided to identify each file.

Click “View/Open” to download.
Simple View versus Full Item Record View

Multiple Files in Simple Item view
- primary file displayed in embedded player
- additional files available for download
Multiple files in Full Item Record View
• Each file displayed in embedded player.
Embedded Audio Player [Example]

mp3 file
Abstract
this is an mp3
View/Open
bullet j23-a-xxx.mp3 (28.47Mb)
Date
2014
Author
Audio, Clip
Metadata
Show full item record

wav file test
Abstract
View/Open
bullet piano2.wav (1.154Mb)
Date
2015-10-28
Metadata
Show full item record

URI
http://hdl.handle.net/123456789/34
Collections
Video Demos
Uploading Audio

- The player supports all file types that HTML5 supports.
- Recommended formats for audio:
  - .mp3, .wav
Uploading Video

• The player supports all file types that HTML5 supports.

• Recommended formats for video:
  • .avi, .mov, .mp4, .asf, .wmv

• It is recommended that the uploaded videos be around 15 frames per second, at anywhere between 320x240 to 1080x720 resolution.
Video and Audio Player Allowed Formats

HTTP 206 + HTML5 audio/video
+HTTP 206 implementation for BitstreamReader -- needed for html5 video/audio players

+Adds html5 player to xmlui/Mirage2 for video files with certain file extensions:
  .3g2, .3gp, .asf, .avi, .m4v, .mov, .qt, .mp4, .m4p, .m4v, .mpg, .mpeg, .m2v, .vob -- as <video> with <source src=xxx type=video/mp4>
  .ogg, .ogv -- as <video> with <source src=xxx type=video/ogg>
  .webm -- as <video> with <source src=xxx type=video/webm>
  .drc, .flv, .mkv, .mng, .mp2, .mpe, .mpv, .mxf, .nsv, .rm, .rmvb, .roq, .svi, .wmv, .yuv -- as <video> with <source src=xxx> with no type

+Adds html5 player to xmlui/Mirage2 for audio files with certain file extensions:
  .weba -- as <audio> with <source src=xxx type=audio/webm>
  .mp3 -- as <audio> with <source src=xxx type=audio/mpeg>
  .oga, .opus -- as <audio> with <source src=xxx type=audio/ogg>
  .wav, .wave -- as <audio> with <source src=xxx type=audio/wav>
  .act, .aiff, .aac, .amr, .ape, .au, .awb, .dct, .dss, .dvf, .flac, .gsm, .iklax, .ivs, .m4a, .mmf, .mpc, .msv, .ra, .raw, .sln, .tta, .vox, .wma, .wv -- as <audio> with <source src=xxx> with no type
Batch Uploading Audio/Video Files

• Batch uploading audio and video files is possible in DSpace 5.
• Depending on file sizes, batch uploads could be a lengthy process.
• We recommend scheduling time for batch uploads accordingly.
Minor Issues and Next Steps

• The video player has issues supporting all MIME types at the moment.
• For videos with multiple parts, only the first part will load in the video player, subsequent parts will open in an installed video player on the computer.
• The programming team is planning enhancements to the video player later this year.
Google Analytics from DSpace Admin UI

LE YANG
Google Analytics from DSpace Admin UI

• If you are logged in, the “View Google Analytics Statistics” feature is visible in the side navigation menu (bottom right)

• This feature pulls statistics of page views and file downloads through Google Analytics
Google Analytics from DSpace Admin UI

• Administrators can select a time range for viewing page views and file downloads

• Page views and file downloads are grouped by month in the display
Google Analytics: File Downloads

• “File Downloads” gives actual usage of each file (i.e., item downloads)

• For example: One file may have 200 page views, but only 20 file downloads. “File Downloads” gives a more accurate view of usage at the item level.

<table>
<thead>
<tr>
<th>File Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015 / 11</td>
</tr>
<tr>
<td>2015 / 10</td>
</tr>
<tr>
<td>2015 / 09</td>
</tr>
<tr>
<td>2015 / 08</td>
</tr>
<tr>
<td>2015 / 07</td>
</tr>
<tr>
<td>2015 / 06</td>
</tr>
<tr>
<td>2015 / 05</td>
</tr>
<tr>
<td>2015 / 04</td>
</tr>
</tbody>
</table>
Batch Upload via DSpace UI

LAURA MCELFRESH AND ROBERT LUTTRELL
Batch Upload via DSpace UI

• Previous versions of DSpace required command-line actions

• DSpace 5 batch upload process is more user-friendly (i.e., graphical interface)

• Requires file/directory arrangement, as in previous versions
Batch Upload via DSpace UI

DSpace 5 Batch Import Format

• Instructions for Standard Archival Format (SAF) packaging:

Batch Import Steps

• Specific Instructions
  ▪ Create and arrange files, zip, and upload
Batch Upload via DSpace UI

Each Item in a batch upload needs a folder containing the following files:

- Item bitstreams
- Metadata
- License
- Contents
Batch Upload via DSpace UI

The “contents” file is a tab-delimited text file.

• You can create it in Notepad, then delete .txt extension from the filename.

• **Note:** If you don’t delete .txt, DSpace will not recognize the file!

• Example: Here’s the contents file for **item1** in previous screenshots, a collection of pictures of flowers:

<table>
<thead>
<tr>
<th>File Name</th>
<th>Bundle Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>license.txt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrysanthemum.jpg</td>
<td>bundle:ORIGINAL</td>
<td>description:Chrysanthemum</td>
</tr>
<tr>
<td>Hydrangeas.jpg</td>
<td>bundle:ORIGINAL</td>
<td>description:Hydrangeas</td>
</tr>
<tr>
<td>Tulips.jpg</td>
<td>bundle:ORIGINAL</td>
<td>description:Tulips</td>
</tr>
</tbody>
</table>
Batch Imports from the User Interface

- DSpace 5 has the ability to import items (photos, PDFs, etc.) AND metadata directly from the Graphical User Interface from a DSpace Administrator user from a zip file in the following format:

  It must follow this format:

  archive_directory/
  item_000/
  - dublin_core.xml - qualified Dublin Core metadata for metadata fields belonging to the dc schema
  - metadata_[prefix].xml - metadata in another schema (the prefix is the name of the schema as registered with the metadata registry)
  - contents - text file containing one line per filename
  - collections - text file that contains the handles of the collections the item will belong two. Optional. Each handle in a row.
  - file_1.doc - Collection in first line will be the owning collection
  - file_2.pdf - files to be added as bitstreams to the item
  - file_1.png
  ...

See below for additional documentation:
https://wiki.duraspace.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format
Upload your file

Select a collection
Collection:
Select the collection you wish to submit an item to.
Special Collections > Postcards > Cape Cod Postcards

Choose File ingest_jdame_20150106-153342.zip

Upload SimpleArchiveFormat ZIP
Batch Upload in DSpace 5

• Recent article in DLIB Magazine:
  • Desktop Batch Import Workflow for Ingesting Heterogeneous Collections: A Case Study with DSpace 5
    http://www.dlib.org/dlib/january16/nash/01nash.html

• TDL is currently working on testing of batch import scripts and will have more updates in the coming weeks
ORCID Lookup

ALEXANDRA MITCHELL
ORCID Lookup

- ORCID is a unique ID that researchers/other users can sign up for – utilized as a name disambiguation system
- DSpace 5 includes an “ORCID Lookup” feature for author and editor fields in Dublin Core
- The “ORCID Lookup” feature is not required and you are still able to add authors and editors without an ORCID profile (as in previous versions)
ORCID System Interface

The ORCID system features options for adding profile information that enables others to correctly identify a researcher with their work.
ORCID Lookup

Author and editor names can still be added by typing in the name and clicking “add” (as in previous versions of DSpace)
ORCID Lookup Screen

Names of people already listed in your repository are displayed (even if they don’t have an ORCID associated with their name)
ORCID Lookup Screen

Names of people with an ORCID account, located by the “ORCID Lookup” feature, will have their name displayed in italics.
Lookup Screen

This is an example of the screen display if you are entering names and there are multiple people with the same name in the ORCID database.

Note: Make sure that you are selecting the correct person.
If you choose to utilize the “ORCID Lookup” feature, be sure that you have selected the correct person.

Clicking the ORCID number through the “ORCID Lookup” feature in DSpace 5 will generate a pop-up window for viewing profile information supplied by the person in ORCID, which may be helpful for ensuring that you are selecting the correct name (i.e., author or editor) associated with the work.
SHERPA/RoMEO

ALEXANDRA MITCHELL
SHERPA/RoMEO

- SHERPA/RoMEO is a database of publishers’ policies “regarding the self-archiving of journal articles on the web and in Open Access repositories”
- Note: You must use the ISSN for the SHERPA/RoMEO policies to appear
- A color coded system is used to identify the publishers’ policies
- Note: If you submit more than 500 requests per day from one IP address, you will need to register for a free account
In order for the “Sherpa/RoMEO Lookup” feature to work in DSpace 5, you must enter the ISSN number of the journal that the article was published in.
SHERPA/RoMEO

This is an example of a journal found through the “SHERPA/RoMEO” Lookup feature.

You can click “More Info” for details about the color coding system.
SHERPA/RoMEO

The color-coded system through SHERPA/RoMEO lets you know the permissions and restriction policies.

If journal information is not available, you will get a “grey” color indication.
When you click “More Info” for a journal that has been found by SHERPA/RoMEO, you will retrieve additional details about that journal’s publishing policies.
MathJax

LAURA MCELFRESH
MathJax: Math and Science Typesetting

MathJax Javascript facilitates the display of mathematical and scientific formulas in DSpace 5.

• You must provide the MathJax coding during the submission process (i.e., DSpace does not perform conversions for you (from MathML, TeX, etc.)

• Note: MathJax code is wrapped with dollar signs
MathJax: Math and Science Typesetting

Sources for MathJax markup how-to:

- [http://www.onemathematicalcat.org/MathJaxDocumentation/TeXSyntax.htm](http://www.onemathematicalcat.org/MathJaxDocumentation/TeXSyntax.htm)
- Have not (yet) found a conversion engine online
MathJax: Math and Science Typesetting

Marked-up scientific/mathematical text will display in multiple fields. For example:

- Author field
- Abstract field
- Title field

| dc.author | Duck, $\lim_{x \to Donald}$ |
| dc.description.abstract | $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ $\require{cancel} \frac{(x+1) \cancel{(x+2)}}{3 \cancel{(x+2)}}$ |
| dc.title | Mathmagics Land: A Travelogue, Vol. $\sum_{n=1}^{\infty} \frac{1}{2^n}$ |

This is how to do MathJax equations using the new DSpace feature documented here https://github.com/DSpace/DSpace/pull/680

THE $1^{1}B^{+}_{u} \leftarrow 1^{1}A^{-}_{g}$ ABSORPTION OF JET-COOLED TRANS,TRANS-1,3,5,7-OCTATETRAENE
MathJax: Math and Science Typesetting
Additional Features and Bug Fixes

ROBERT LUTTRELL
NERISSA LINDSEY
SUSAN ELKINS
OAI-PMH and OpenAIRE

The OAI-PMH Interface may be used by other systems to harvest metadata records from your DSpace. There have been several bug fixes (including the QDC list records issue seen in DSpace 4.x) to the OAI data so they should now follow the correct OAI standard.

Here is the link for the DSpace 5.x Documentation on OAI PMH Harvesting
https://wiki.duraspace.org/display/DSDOC5x/OAI

To View your OAI data from your DSpace server input the following URL into the browser, specifying your specific DSpace server information:
https://YOURSERVER.tdl.org/YOURSERVER-oai/request?verb=Identify

Example:
https://dspace-test.tdl.org/test-oai/request?verb=Identify

To see if you server is validated on the OAI standards go to:
http://validator.oaipmh.com/

Put in your OAI address for validation, and run the validation check ie:
https://dspace-test.tdl.org/test-oai/request

This version of DSpace has been updated for the European OpenAIRE standard as well. TDL has verified that its test server has passed OpenAIRE’s verification server.
PDF Thumbnail Generation

- In DSpace 5, PDF thumbnails are generated from the first page of the file uploaded (or primary file).
- Thumbnail images are generated nightly, when the DSpace media filter process is run – in conjunction with OCR bitstreams and updates of the search index.
- Thumbnails generated are now “clickable,” generating a download of the primary file.
PDF Thumbnail Generation
Page Not Found Bug Fix

Old version

When a user selects a page that is no longer there, a 404 Error is returned. In previous versions of DSpace the page was a simple statement of “Resource not found” that looked like a more serious error. This bug fix now shows “Page cannot be found. Go to DSpace home. Please contact the site administrator if you wish to report this error. If possible, please provide details about what you were doing at the time this error occurred.”
Performance Improvements

Performance improvements for “Select Collection” dropdown in submission process.

• The “select collection” step in the DSpace workflow had a bug that caused it to perform badly when a large number of collections existed

• As per the documentation: https://jira.duraspace.org/browse/DS-682 this bug has been fixed
Questions and Discussion

TDL Helpdesk: support@tdl.org