

Intro to DSpace



**AUGUST 19, 2009
BAYLOR UNIVERSITY**

**CAMERON KAINERSTORFER
TECH SUPPORT SPECIALIST
CAMERON.KAINERSTORFER@UTSOUTHWESTERN.EDU**

**HEATHER PERKINS
METADATA MANAGER
HEATHER.PERKINS@UTSOUTHWESTERN.EDU**

Schedule



- 9:00—Introduction to DSpace
 - 9:30—Users, Groups, and Authorizations
 - 10:00—Hands-on: Creating Collections
 - 10:30—Item Submission and Workflow
 - 11:00—Hands-on: Submitting an Item
- Lunch 11:30 – 1:00*
- 1:00—Metadata Registries and Templates
 - 1:30—Administering DSpace
 - 2:00—Configuration Options
 - 2:30—Batch Imports

Introduction to DSpace



What is DSpace?



- An open source software package that provides the tools for management of digital assets.
 - ✧ *CAPTURES*
 - ✧ *DESCRIBES*
 - ✧ *DISTRIBUTES*
 - ✧ *PRESERVES*

CAPTURES...



- Digital material in any format
- If desired, directly from creators (faculty, etc.)
- Large-scale, stable, managed long-term storage

DESCRIBES...



- **Metadata**
 - Descriptive
 - Technical
 - Rights
- **Persistent identifiers**
 - “handles”

DISTRIBUTES...



- **Via WWW, with necessary access control**
 - [ResearchWorks at the University of Washington](#)
 - [Loughborough University Institutional Repository](#)

PRESERVES...



- **Bitstream guaranteed**

Who is using DSpace?



- Over 400 registered institutions worldwide
 - [List of United States users available online](#)
- More than 1 million digital assets; largest sites contain several hundred thousand items
- Primarily research/higher education institutions
- Cultural heritage organizations, state libraries/archives
- Some commercial users and service providers
- Active development community

DSpace—Then and Now



- 2002—came out of a partnership between developers at Hewlett-Packard and MIT
- 2007—DSpace Foundation (non-profit organization) formed
- 2009—Announcement of DuraSpace
 - A merger of Fedora Commons and DSpace Foundation
 - The two flagship repository platforms will be sustained
 - Will offer new technologies and services responding to the dynamic environment of the Web and to new requirements from existing and future users

DSpace and TDL (Texas Digital Library)



- 2005—TDL formed by four Texas members of the Association of Research Libraries
- Now includes [18 member institutions](#)

Resources—DSpace

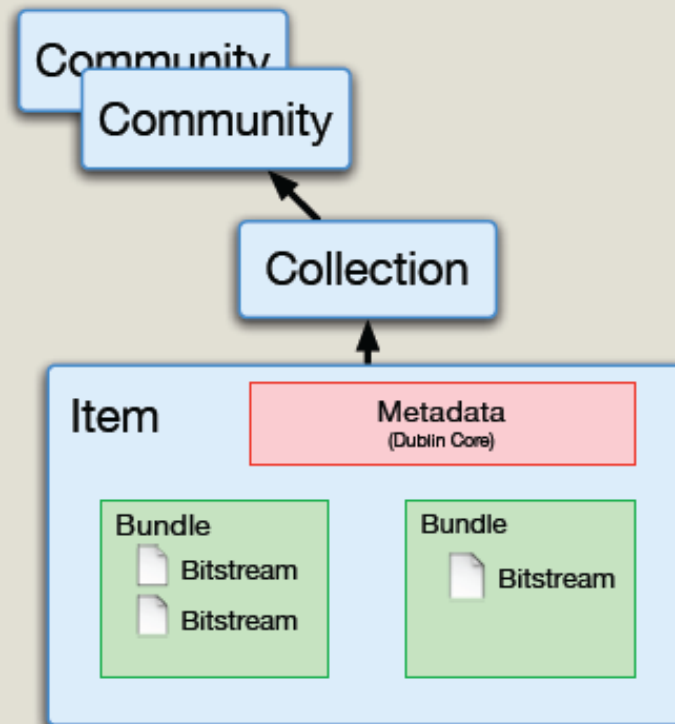


- <http://www.dspace.org/>
 - Documentation
 - [NewSpace Newsletter](#)
- **Mailing Lists**
 - Dspace-general
 - dspace-tech
 - dspace-devel
- <http://wiki.dspace.org/>
 - Technical guides
 - Ongoing projects

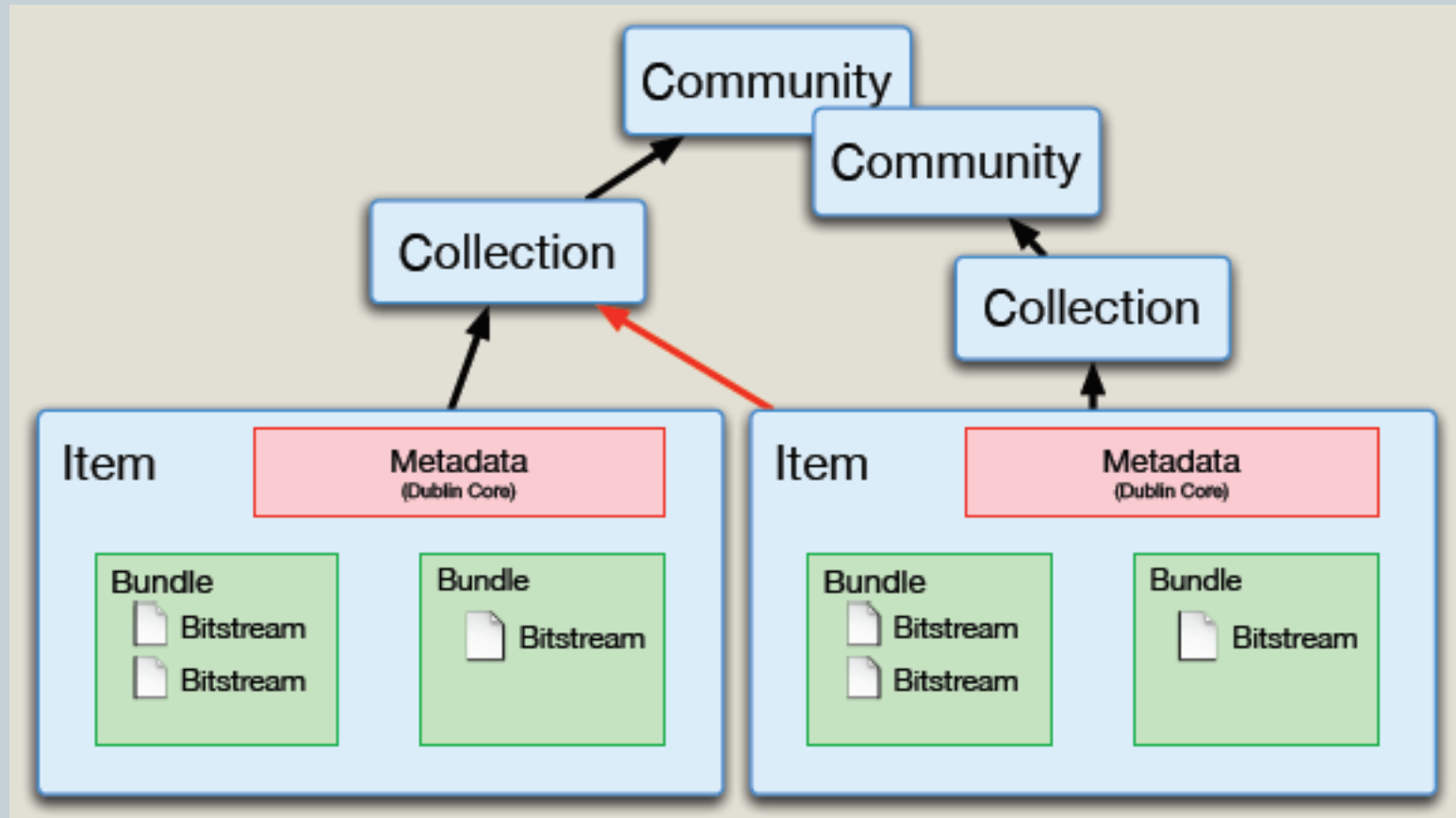
DSpace Data Model



DSpace Data Model



DSpace Data Model



Metadata



- **Descriptive Metadata**
 - Qualified Dublin Core
 - Limited expansion to other formats
- **Administrative Metadata**
 - Internal access control—who can access something for changes
- **Structural Metadata**
 - Bundles & bitstreams—describes location, what belongs to an item

Handles



- **Persistent Identifier—globally unique—attached to objects**
 - Communities
 - Collections
 - Items
 - **Format—can be written in two forms**
 - (as identifier) Hdl: 1721.123/4567
 - (Web browser can resolve this) <http://hdl.handle.net/1721.123/4567>
- ✦ Prefix
 - ✦ Identifier

External Linking



- **Repository item**

- Handle

- ✧ <http://repository.tamu.edu/handle/1969.1/6885>

- **Bitstream**

- ✧ Handle + Bitstream Name

- ✧ <http://repository.tamu.edu/handle/1969.1/6885/ESL-HH-86-11-25.pdf>

Communities and Collections



- **Communities**
 - Can contain sub-communities OR collections
- **Collections**
 - Can contain items
- **Items**
 - Contains metadata and bitstreams (files)

Example (“Bare Bones”)



- **Community**
 - **Sub-community**
 - ✦ **Sub-community**
 - **Collection**
 - **Items**
 - ✦ **Collection**
 - **Items**
 - **Collection**
 - ✦ **Items**

Example (“More Recognizable”)



- **University (Community)**
 - College (Sub-community)
 - ✦ Department (Sub-community)
 - Faculty Member (Collection)
 - Photographs (Items)
 - ✦ Center (Collection)
 - Technical Reports (Items)
 - Historical Images (Collection)
 - ✦ Images (Items)

Example (“Repository”)



- **UT Southwestern (Community)**
 - **Department of Clinical Sciences (Sub-community)**
 - ✦ **Biostatistics Division (Sub-community)**
 - **Philip Baumgartner, MD (Collection)**
 - **Datasets (Items)**
 - ✦ **UT Southwestern Library (Sub-community)**
 - **Photographic History of the UT Southwestern Campus (Collection)**
 - **Photographs (Items)**
 - **Newsletter 1970 – 1976 (Collection)**
 - **PDFs (Items)**

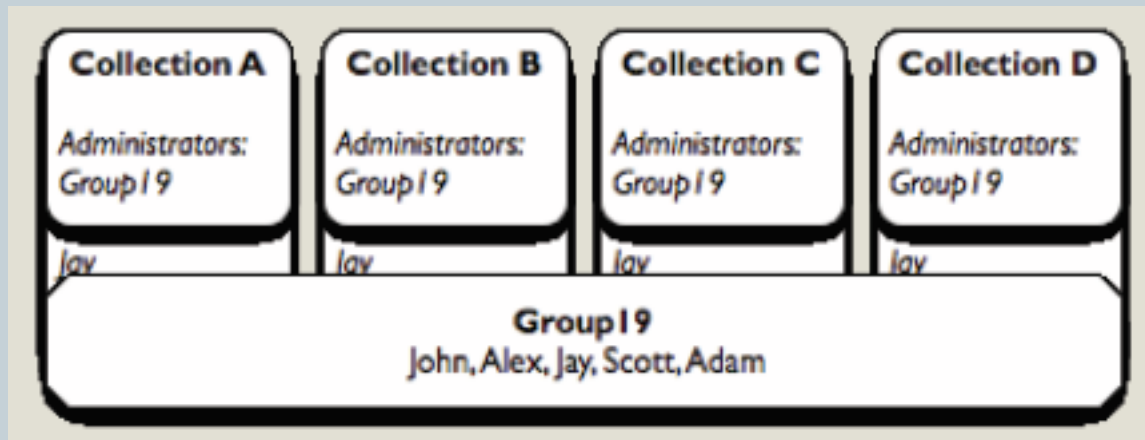
Users, Groups, and Authorizations



Users and Groups



- **Users**
 - Individuals that have a role in the system
- **Groups**
 - Groups of users that share roles



Special Groups



- **“Administrator”**
 - Group of system administrators
- **“Anonymous”**
 - Anyone

User Authorizations



- **Bitstream**

- READ – can open the file
- WRITE – can alter the file

- **Bundle**

- ADD/REMOVE – can add bitstreams to a bundle

- **Item**

- READ – can view the item
- WRITE – can modify the item
- ADD/REMOVE – can add or remove bitstreams

- **Collection**

- ADD/REMOVE – can add or remove items from the collection
- DEFAULT_ITEM_READ – new items receive this READ attribute
- DEFAULT_BITSTREAM-READ – new bitstreams receive this READ authorization
- COLLECTION_ADMIN – can edit or withdraw items, or map items into the collection

- **Community**

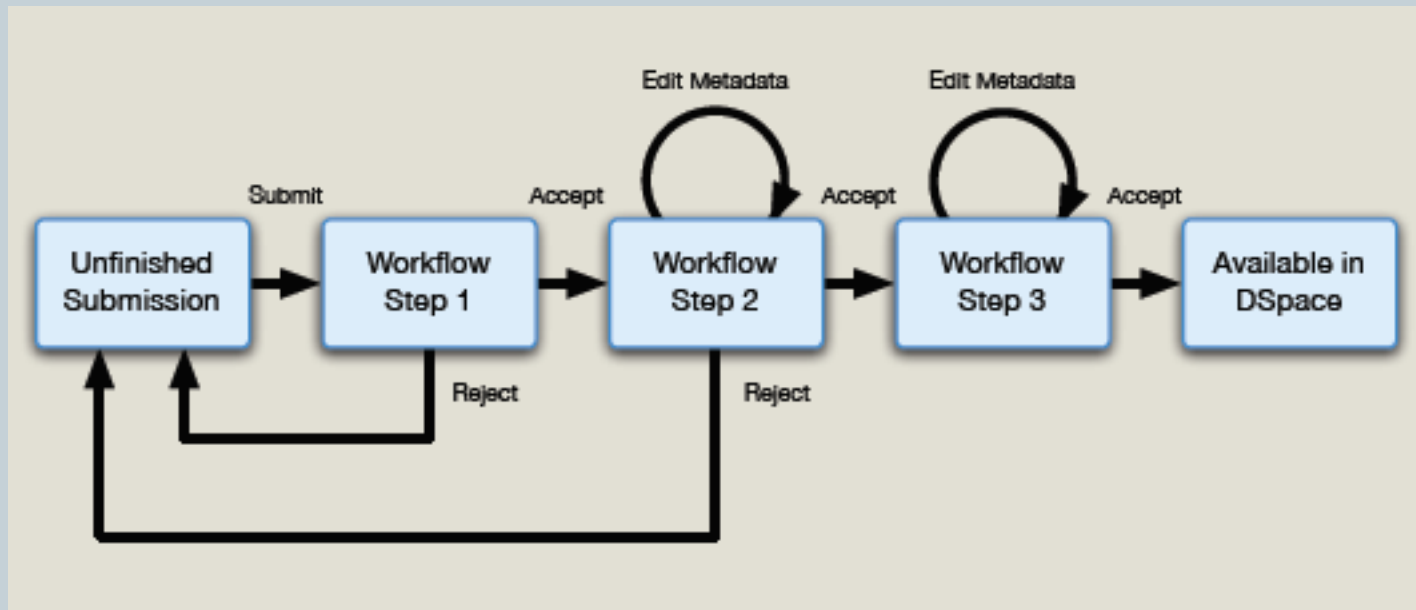
- ADD/REMOVE – can add or remove collections from the community

Workflow Steps



- **No Workflow Steps**
 - Item is made available upon submission
- **Workflow Step 1**
 - Administrator can accept or reject a submission
- **Workflow Step 3**
 - Administrator can edit metadata before making item available
- **Workflow Step 2**
 - Combination of 1 and 3

Workflow Steps



Hands-On: Communities and Collections



- **Creating Communities**
 - Please use your initials for your first community
- **Creating Collections**
 - Folders available on desktop to upload for images, text files, logos, etcetera

Submission and Approval



- **Submission** (web-based, single item)
 - Metadata entry
 - File upload
 - License agreement
- **Approval**
 - Depends on the activated workflow step
 - *Accept/Reject*
 - ✦ Accept puts the item into public view
 - ✦ Reject sends the workflow back to the submitter
 - *Edit Metadata*
 - ✦ Once metadata is edited, item goes into public view

Hands-on: Submitting an Item



Hands-on: Workflow with Items



Metadata Registries and Templates



- **Metadata Registry**

- Defines metadata fields
- Add new fields

- **Item Template**

- Set default values for metadata fields
- Affects all new submissions, does not change metadata for existing items

Item Mapper



- Items can appear in multiple collections
- Must be mapped from the destination collection
- Mapped items **appear** in the second collection but do not **exist** in the second collection
- *Context* → *Item Mapper* option will only show if there is something to map over to the other collection

Administrrating DSpace



Removing Items: Withdraw vs. Delete



- **Withdraw**
 - Removes item from view
 - Does not show up in search results
 - Recoverable
- **Permanently Delete (“Expunge”)**
 - Unrecoverable
 - Handle is not reused
 - Can only be done by a repository administrator

External Linking



- **Repository item**

- Handle

- ✧ <http://repository.tamu.edu/handle/1969.1/6885>

- **Bitstream**

- ✧ Handle + Bitstream Name

- ✧ <http://repository.tamu.edu/handle/1969.1/6885/ESL-HH-86-11-25.pdf>

System-wide Alert



- Accessed through the Control Panel menu item
- Can be used to notify users of downtime or other maintenance
- Timer can be added to note expiration
- Other possible uses

Configuration Options



Configuration Locations



- **File:** dspace.cfg
 - General DSpace parameters
 - Catch all location
- **File:** xmlui.xconf
 - Where (Manakin) themes are installed
 - Interface plugins, known as aspects
- **Files:** input-forms.xml
 - Configure the submission questions

dspace.cfg: Email Options



From address for mail

mail.from.address = dspace-noreply@myu.edu

Currently limited to one recipient!

feedback.recipient = dspace-help@myu.edu

General site administration (Webmaster) e-mail

mail.admin = dspace-help@myu.edu

#Recipient for server error and alerts

#alert.recipient = email-address-here

#Recipient for new user registration emails

#registration.notify = email-address-here

dspace.cfg: Search & Index



search.index.1 = author:dc.contributor.*
search.index.2 = author:dc.creator.*
search.index.3 = title:dc.title.*
search.index.4 = keyword:dc.subject.*
search.index.5 = abstract:dc.description.abstract
search.index.6 = author:dc.description.statementsresponsibility
search.index.7 = series:dc.relation.ispartofseries
search.index.8 = abstract:dc.description.tableofcontents
search.index.9 = mime:dc.format.mimetype
search.index.10 = sponsor:dc.description.sponsorship
search.index.11 = identifier:dc.identifier.*
search.index.12 = language:dc.language.iso

dspace.cfg: Authentication



- **Password (default)**
 - User's signup for an account with the repository
- **LDAP**
 - Access one university's local account management system
- **Shibboleth**
 - Access multiple universities account management systems

dspace.cfg: Browse Indexes



- **Browse Metadata**
 - Name
 - Metadata fields
 - Data type (title, text, date)
- **Browse Items**
 - Name
 - Sorting option
- **Sorting Options**
 - Name
 - Metadata field
 - Data type (title, text, date)

dspace.cfg: Browse Indexes



Browse Configuration

webui.browse.index.1 = dateissued:item:dateissued
webui.browse.index.2 = author:metadata:dc.contributor.*:text
webui.browse.index.3 = title:item:title
webui.browse.index.4 = subject:metadata:dc.subject.*:text
webui.browse.index.5 = dateaccessioned:item:dateaccessioned

Sorting Options

webui.itemlist.sort-option.1 = title:dc.title:title
webui.itemlist.sort-option.2 = dateissued:dc.date.issued:date
webui.itemlist.sort-option.3 = dateaccessioned:dc.date.accessioned:date

xmlui.xconf: Themes and Aspects



- **Install Manakin Themes**
 - For the whole repository
 - For specific communities
 - For specific collections
 - For a specific page
- **Install Manakin Aspects:**
 - For the whole repository

input-forms.xml: Metadata fields

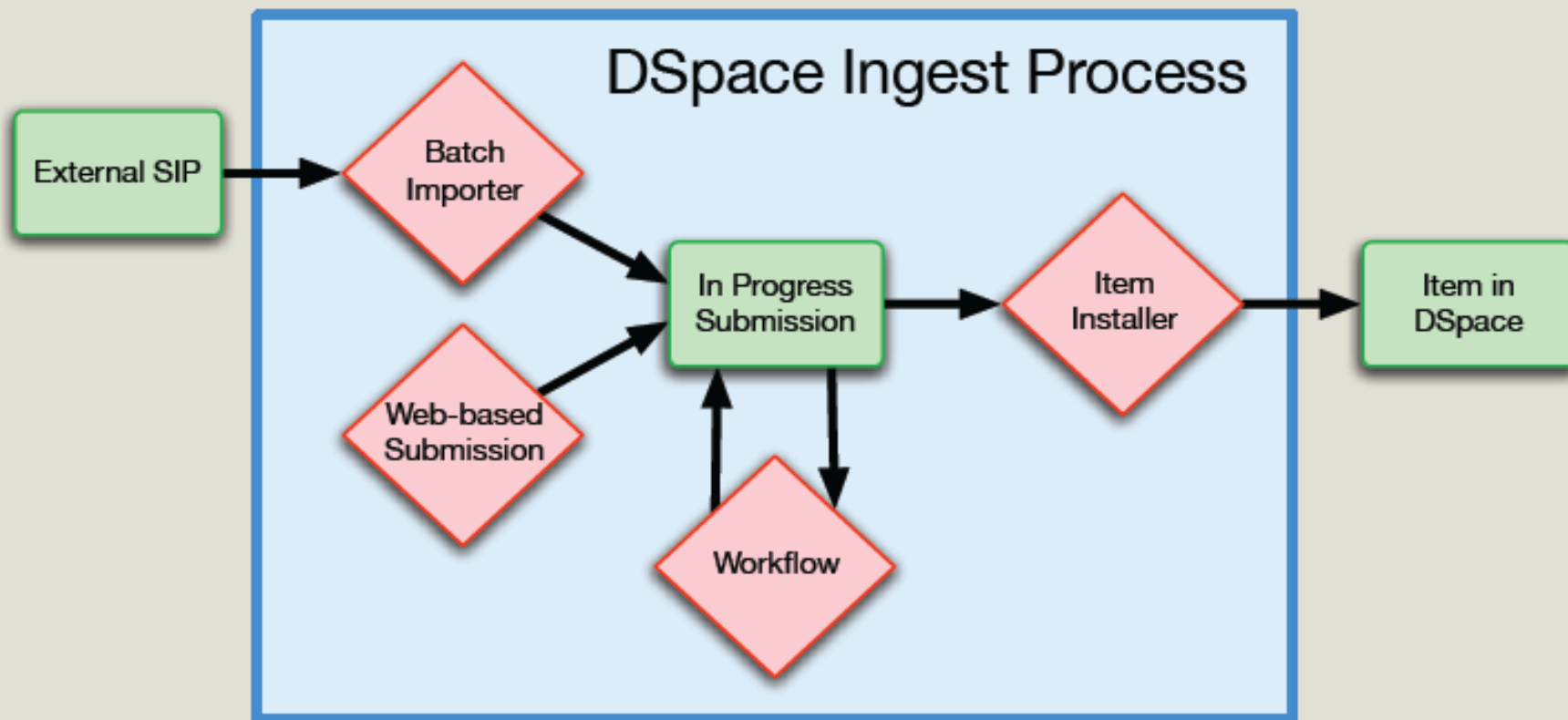


- Define the questions asked during an item's submission
- Create forms that can be attached to particular collections for:
 - How many pages or steps there are to describe an item
 - What metadata fields are presented on each page

Batch Import



Ingest Process



Batch Import



- **Command line import**
- **Directory structure**
 - “contents” file
 - “dublin_core.xml” file
 - “handle” file
 - bitstreams files

Simple Contents File



- **Format:**

`<filename>`

or

`<filename> <tab> bundle:<bundle name>`

- **Examples:**

dissertation.pdf

mods.xml

license.txt

bundle:METADATA

bundle:LICENSE

Dublin Core Metadata



- **Example:**

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<dublin_core>
```

```
    <dcvalue element="contributor"
```

```
    qualifier="author">John</dcvalue>
```

```
    <dcvalue element="language" qualifier="iso">en</dcvalue>
```

```
    <dcvalue element="subject"
```

```
    qualifier="none">Technology</dcvalue>
```

```
    <dcvalue element="title"
```

```
    qualifier="none">Sample Title</dcvalue>
```

```
</dublin_core>
```

Import Command



```
./dsrun org.dspace.app.itemimport.ItemImport
```

- | | |
|------------|--|
| -a | Add new items to DSpace |
| -c <coll> | Which collection to add them to |
| -e <email> | Existing user who is adding these items |
| -m <path> | Create a log file for this import |
| -s <path> | Location of the import files |
| -t | Do not run, just test the import for validity |
| -h | Print command line options and their description |

Import Command (example)



Examples

```
./dsrun org.dspace.app.itemimport.ItemImport -a  
-c 123456789/5 -e scott@library.tamu.edu  
-m /path/to/file.map -s /path/to/import
```

-a	Add new items to DSpace
-c <coll>	Which collection to add them to
-e <email>	Existing user who is adding these items
-m <path>	Create a log file for this import
-s <path>	Location of the import files

More Resources—DSpace (TDL)



- **About the Texas Digital Library**
- **TDL News**
- **TDL Publications**
 - **TDL Demonstration Video**—this is a good short introduction to how an item can be submitted at one institution and someone else at another institution is able to review metadata and complete the submission (available from the TDL Publications area)
 - **DSpace Batch Import Format**—an alternative to single item submission, this explains the batch import system

Upcoming Training Opportunities



- **September 21**
 - DSpace Customization – full day
 - Open Journal Systems – full day

Full list:

<http://www.tdl.org/about-tdl/training/>