Welcome To The Connecticut Digital Archive

This Presentation Covers:

- What is the CTDA?
- What are the technologies of the CTDA?
- Being Responsible Owners of Your Content
- Namespaces
- The Sandbox
- Your Profile
- Top Level Navigation
- Managing content
- Users
- Resources

What is the CTDA?

- Mission:
  - The CTDA provides long-term preservation services for Connecticut based non-profit institutions.
  - Preservation services include
    - Technical infrastructure such as updated hardware and software to preserve your digital content
    - End-user support such as training participants on how to use the software provided and a support desk to answer questions
    - Ability to add/manage content
- Technology:
  - The CTDA is a digital repository built on the Open Archival Information System (OAIS) developed by NASA (ISO 14721:2003).
  - OAIS is a conceptual model of an information object that is self-contained and self-describing. An information object is a set of data elements combined into an internally coherent package that is understood in the electronic and web environment.

The CTDA and You

- The CTDA
  - Offers participating institutions the wherewithal to take responsible ownership for their digital content without having to worry about the technical hardware and software to preserve digital content over the long-term
- Participating Institutions
  - Benefit from the technical infrastructure support, end-user support, deposit, and access services offered at little to no cost
  - Are able to focus on content, data ownership and responsibility
    - Institutions decide:
      - How to describe your content in the repository
      - What gets preserved or not in the repository
      - The organization of content in the repository
      - Workflows to add/manage content in the repository, including how to pre-process content
      - What content gets aggregated in particular sites like Connecticut History Illustrated
      - Who works on projects, etc.

Responsible Ownership

- Creating Workflows that work for your organization
- Developing pre-processes to prepare digital content to be added to the CTDA such as file naming conventions, data entry guidelines for descriptive metadata, where to store your digital content that is to be loaded, which content models to use, or which digital assets to share
- Learning how to add and manage content and associated descriptions (metadata) to the CTDA
- Organizing your digital content your way
    - Collections can be nested or not
    - Digital assets can be “shared” or “migrated”
    - Different content models can be applied to content.
- Maintaining your digital content and associated descriptions (metadata) in the CTDA
- Monitoring and refining your organizational strategy as necessary

Primary Technologies of the CTDA

*Currently the CTDA's technical infrastructure relies on 4 technologies: Fedora, Islandora, Drupal and the Handle System. What is important about these technologies is the **Vocabulary**. Terminology used in Fedora will appear when you are adding and/or managing content. This section introduces technology but focuses on vocabulary.*

- Fedora (http://www.fedora-commons.org/about)
  - Not a database but a conceptual framework
  - Not a software application but a generic repository foundation upon which many kinds of applications can be created
  - Not an integrated solution or information silo but a basis for software systems to manage digital information
- Islandora (http://islandora.ca/about)
  - Administrative, Discovery and Presentation layer build on Drupal
- Handle System (http://www.handle.net/)
  - Persistent and unique identifier generation system and lives in the MODS metadata

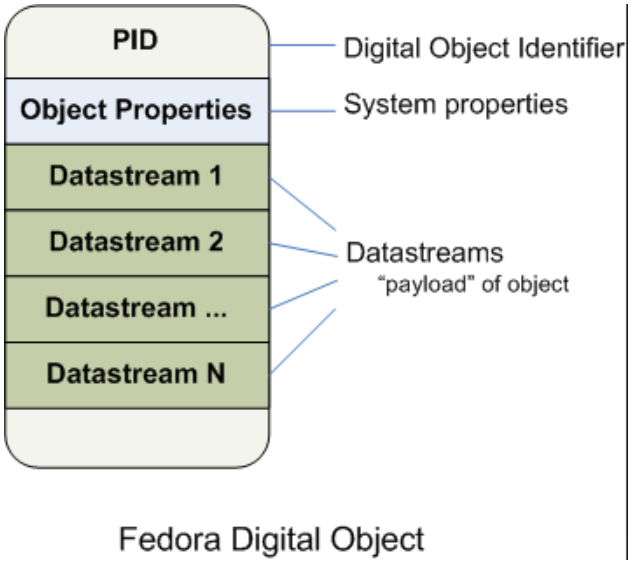
Handle System

The Vocabulary of Fedora

- Fedora, or Flexible Extensible Digital Object Repository Architecture, and relies on the **Digital Object** model.
  - Digital Object: An aggregation of local and distributed information

  The digital object is also referred to as the repository package or packet as it is a stand-alone entity that consists of information about a content item such as a page, book, image, PDF, video, or collection.

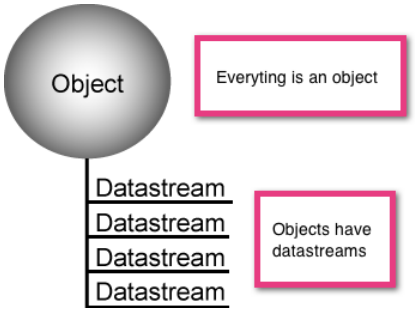
- Digital Objects have:
  - A local persistent identifier called the "**PID**"
  - Object properties such as "coordinates" that tell Fedora where the object belongs in relation to other objects (rels-ext, rels-int)
  - **Datastreams**
    - Information or metadata about the content item such as descriptive, rights, administrative, or technical metadata
    - Information or derivatives for content items such as when you ingest a tiff, the system creates a thumbnail and medium sized jpeg.
    - Information or the content item itself such as the digital asset your ingest (tiff, PDF, mp4, etc.)

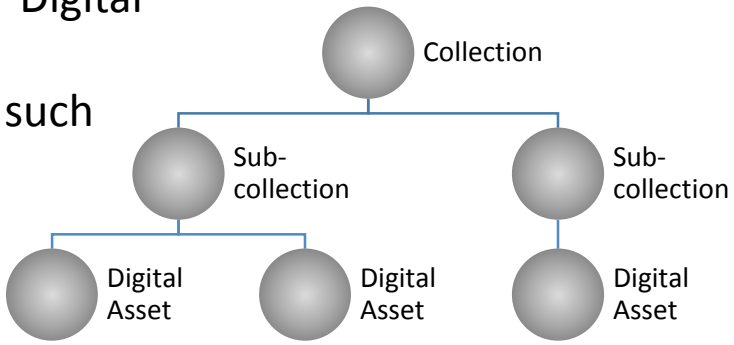


Fedora Digital Object

Everything is an Object!

- CTDA is an atomic universe of Objects whether or not that Object is directly related to a content item such as photograph, document, or video
- All Objects are related thanks to the Fedora Digital Object Properties
- CTDA has only 2 types of Fedora Digital Objects
  - Collection – used to aggregate other Fedora Digital Objects such as other Collections or Digital Asset Objects
  - Digital Asset – used to store a digital asset such as a tiff, jpeg, PDF, etc. (your actual digital content)





Key Vocabulary From Fedora

- **Digital Object or Object**: This refers to the Fedora Digital Object which is the information packet that consists of a set of streams, an object identifier, and object properties
- **Datastream or data stream**: Information about a Fedora Digital Object such as MODS (Metadata Object Descriptive Schema) data stream, DC (Dublin Core) data stream, TN (Thumbnail) data stream
- **Object properties**: Information about how the Fedora Digital Object relates to other objects such as rels-ext (External Relationships, coordinates), rels-int (Internal Relationships, books), how or when users can access the Object, who owns the object, collection policies (what format can be put into a collection), audit information
- **PID**: Fedora local identifier for an Object
- **Collection**: A Fedora Digital Object that acts as a container, like a folder, that groups together a set of other Digital Objects (either other collections or digital assets)
- **Digital Asset**: A Fedora Digital Object that consists of information and the primary digital asset (content) such as the tiff, PDF, video, etc.

Islandora Terminology

- **Drupal modules**: Islandora is a framework or extension to Drupal, a content management system. Drupal uses modules as one way to provide instructions on how to add, manage, and display content.
- **Administrative layer**: Islandora works in tandem with Drupal to provide the tools to add and manage digital content.
- **Presentation layer**: Islandora works in tandem with Drupal to present repository content on the web.
- **Multisite implementation**: The CTDA has multiple Islandora/Drupal sites designed for specific purposes such as to only present repository content, present content related to only Connecticut, or to add/manage content.

Manage CTDA Collections

Connecticut State Library

UNIVERSITY LIBRARIES
Archives & Special Collections at the Thomas J. Dodd Research Center

Trinity College Library

The Connecticut Digital Archive Collections

CTDA Vocabulary Inherited from Islandora

The CTDA is made up of several partners who all have unique content but share the same digital repository. The CTDA uses **NAMESPACES** or a series of digits to identify content from an institution and sometimes a particular set of content within that institution. One institution can have 1+ namespaces.

For a list of our most current namespaces, see http://ctdigitalarchive.org/about/technology/namespaces/.

Islandora Sees the World Using Fedora Vocabulary

- ***Content Models*** or ***Islandora Content Model Digital Objects*** determine how repository Fedora Digital Objects are created and managed in the repository and displayed to the user.
- They are **for the most part** associated with specific file formats because these Content Models tell the system if and which…
  - Viewer to call
  - Derivatives to create
  - Derivatives that are available for download
  - Technical metadata is extracted
  - Object properties are created
  - Specific relationships to other Fedora objects to be created
  - Specific access policies
  - *And so much more…*

Key Vocabulary from Islandora

General Content Models

- **Collection**: Used to create a collection (no format as no content is involved)
- **Large Image**: tiff, jp2 (Viewer has zoom functionality)
- **Basic Image**: jpg, png, bmp, gif (Simple viewer with no zoom functionality)
- **PDF**: PDF files (PDF.js viewer replicates PDF viewer for user in display)
- **Audio**: wav, mp3
- **Video**: ogg, mp4, mov, qt, m4v, avi
- **Book and page**: A specialized set of content models for ordered sets of digital files. The book is a container. Pages need to be tiff/tif or jp2 format. There is also a book batch import that can take a PDF file. Uses the Internet Archive Viewer
- **Newspaper, issue and page**: A specialized set of content models for serial publications or regularly recurring ordered sets of digital files. The newspaper and issue are containers. Pages need to be tiff/tif, jp2, or jpg format. There is also a batch import that can take a PDF. Uses the Internet Archive Viewer for Issues
- **Compound Object**: A specialized set of content models for sets of digital files that combine other Objects in an explicit relationship (i.e. front and back of a postcard). All child objects have to be ingested but can be “hidden”.

General Content Models

- **Document:** Used for common documents (odt, sxw, rtf, doc, wpd, txt, html, ods, sxc, xls, csv, tsv, odp, sxi, ppt, pdf). Batch import will work for all formats**.**
- **WebARChive:** Used for the web archive format (warc), a format revision of the Internet Archive's ARC file format (warc). Upon ingest can also include a screen shot of the web site (as png or jpg) and accompanying txt file.
- **Finding Aid:** Used to add archival finding aids (EAD xml files)
- **Binary Object**: A specialized content model for alternative formats not listed above

Institutional Repository Content Models
For Scholarly Output

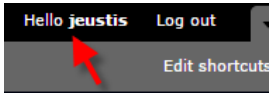
- **Thesis:** Used for theses and dissertations. This includes the ability to embargo objects or datastreams.
- **Citation:** A specialized content model that will record a bibliographic citation according to a specific citation style guide such as APA or Chicago Manual of Style. Batch import can be used for PubMed articles. This is particularly useful for faculty citations to materials that cannot be added to a publicly available repository.
- **Entity/Person:** Used to add an object about a person such as a profile of a faculty member. This doesn't involve content and is a specialized container object – specialized in terms of how it is displayed and the metadata used to describe the person (EAC-CPF or MADS).
- **Entity/Place:** Used to add an object about a place such as a profile of a city. This doesn't involve content and is a specialized container object – specialized in terms of how it is displayed and the metadata used to describe the person (EAC-CPF or MADS).
- **Entity/Organization:** Used to add an object about an organization such as a profile of a vessel. This doesn't involve content and is a specialized container object – specialized in terms of how it is displayed and the metadata used to describe the person (EAC-CPF or MADS).
- **Entity/Event:** Used to add an object about an such as a profile of a hurricane. This doesn't involve content and is a specialized container object – specialized in terms of how it is displayed and the metadata used to describe the person (EAC-CPF or MADS).

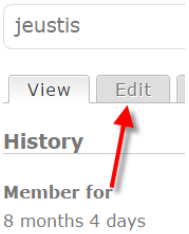
CTDA Sandbox

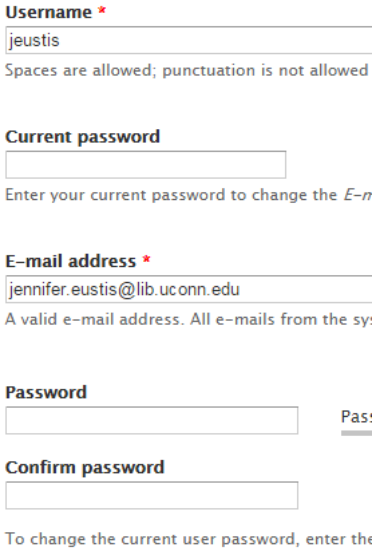
http://ctda-stage.lib.uconn.edu/

Your Profile

- You can access your profile in one of 2 ways.
  - After you log in
  - By clicking on your username in the upper right corner
- To make changes to your profile, click "Edit"
  - You can change your username and password
  - You can also upload a profile picture
- Remember to save all changes
  - Click on "Save" at the bottom
- If you forget your password, just email ctda@uconn.edu

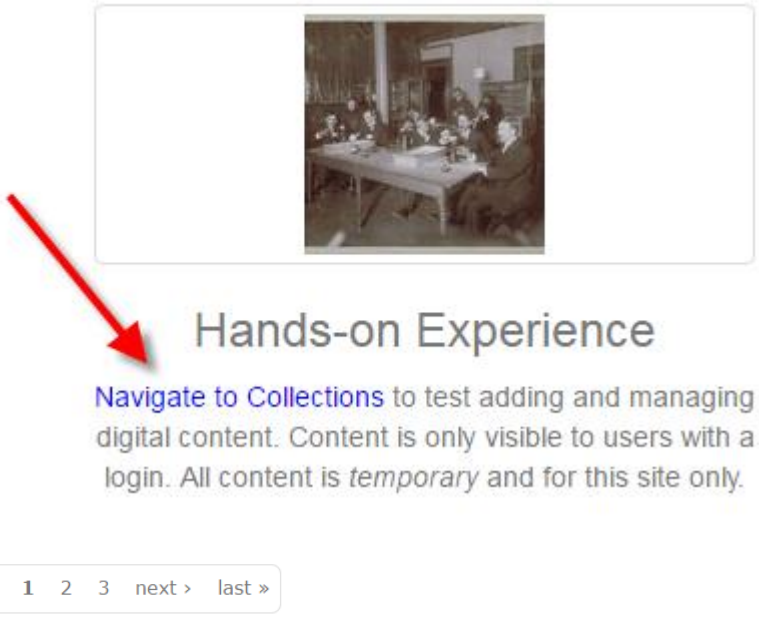






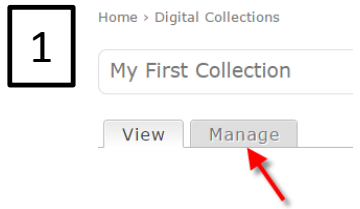
Top Level Navigation

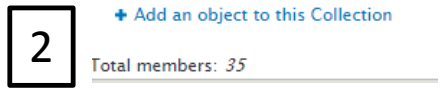
- Getting to Collections
  - Click on “CTDA Stage”
  - Or search for something
- If you click on CTDA Stage, to see all the test collections, click on “Test Collections”
- You can use the “next”, “last”, etc. on the right hand side to navigate all the pages of Collection Objects (for the most part these will be Collection Objects and not Digital Asset Objects)
- To see what’s “in” a collection, just click on it.

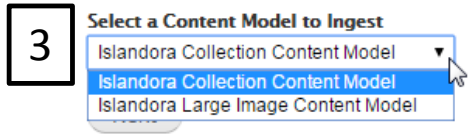


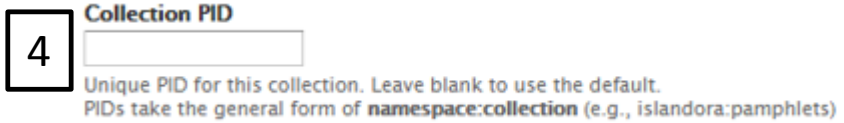
Add A New Collection (Object)

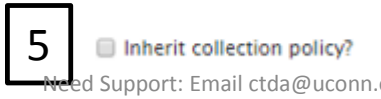
- No matter where you are in the system, it is the same process to create a Collection Object. How does CTDA know where to create your Collection? ***You have to go to that location and then click manage***.
- Create a collection
  - Go to where you want it to be created (top level or subcollection)
  - Click Manage (1)
  - Click "Add an object to this Collection" (2)
  - Select "Islandora Collection Content Model" (3)
  - Click Next
  - Fill out the Collection PID (4)
  - ***UNCHECK INHERIT COLLECTION POLICY*** (5)





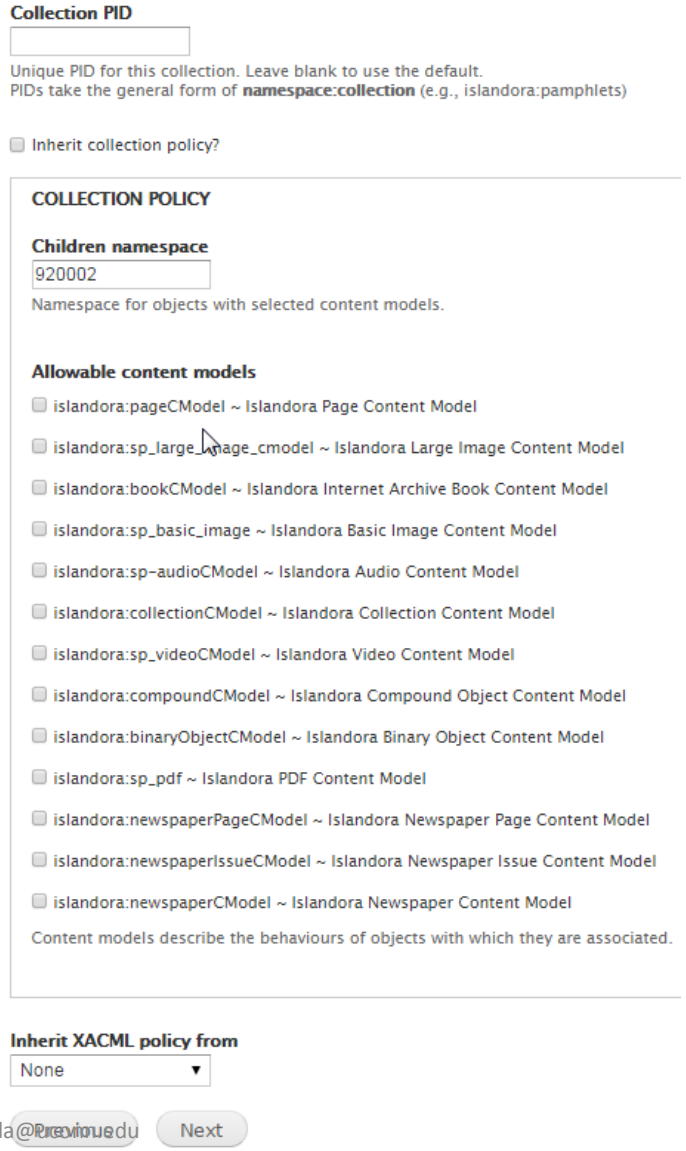






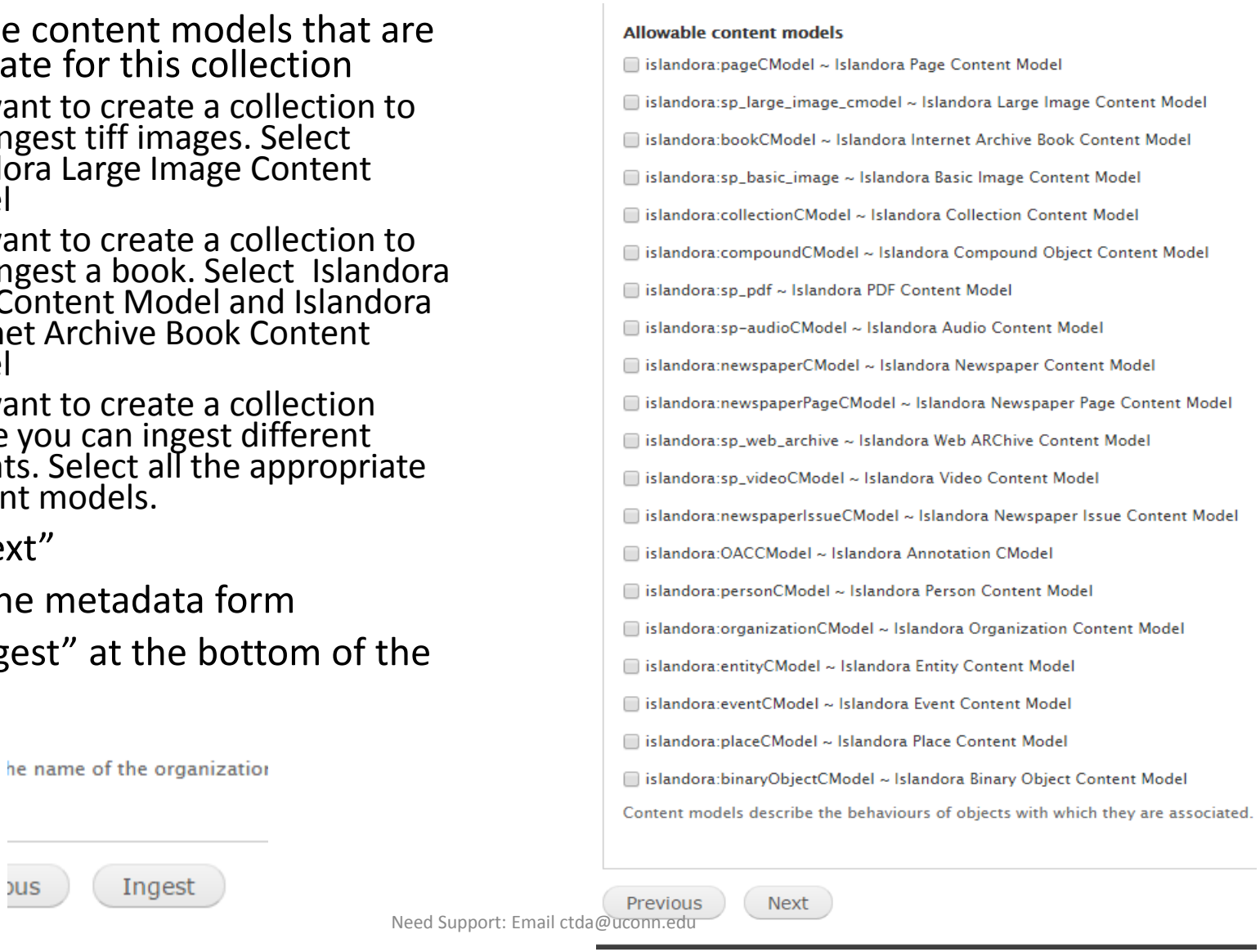
Collection

- The Collection PID **MUST** have ***your institution's namespace colon and then any alphanumeric string***
  - PIDs cannot contain any spaces or unusual characters such as slashes, question marks or brackets
  - PIDs can contain underscores and hyphens
- **UNCHECK** "Inherit collection policy"
- Ignore "Inherit XACML policy from"



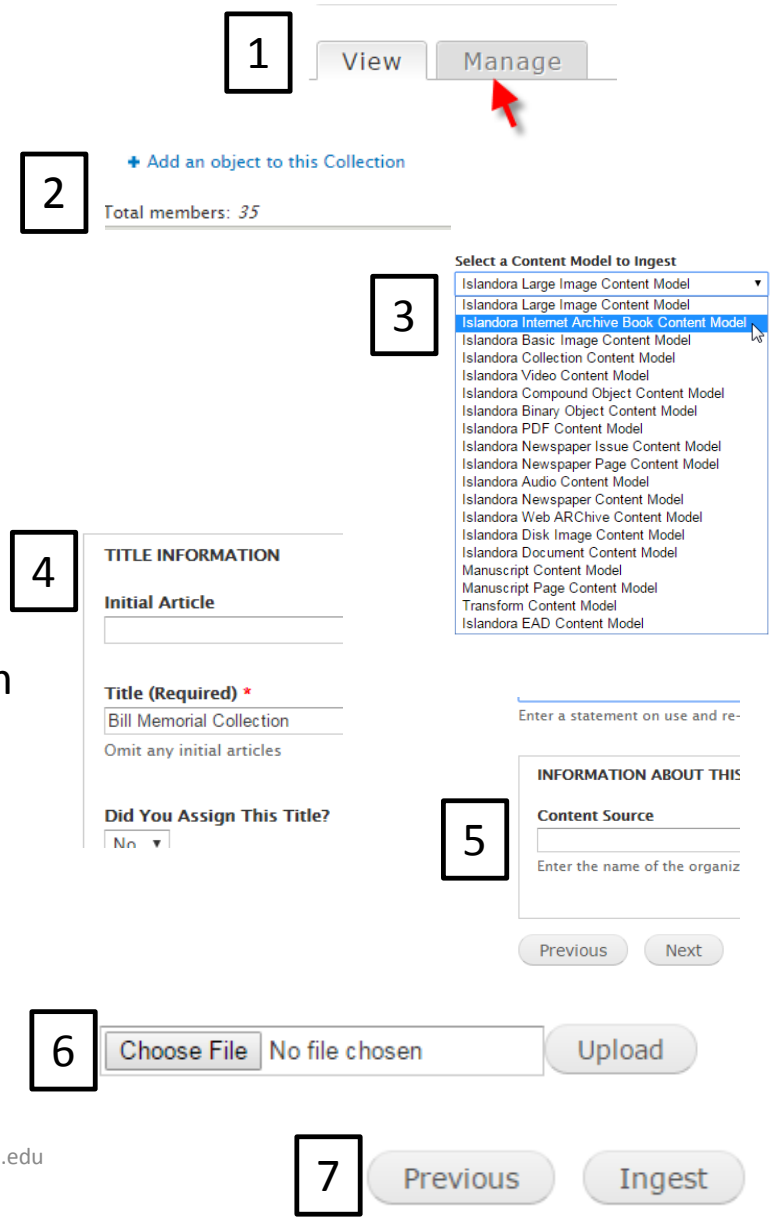
Collection

- Select the content models that are appropriate for this collection
  - You want to create a collection to only ingest tiff images. Select Islandora Large Image Content Model
  - You want to create a collection to only ingest a book. Select Islandora Page Content Model and Islandora Internet Archive Book Content Model
  - You want to create a collection where you can ingest different formats. Select all the appropriate content models.
- Click “Next”
- Fill out the metadata form
- Click “Ingest” at the bottom of the form



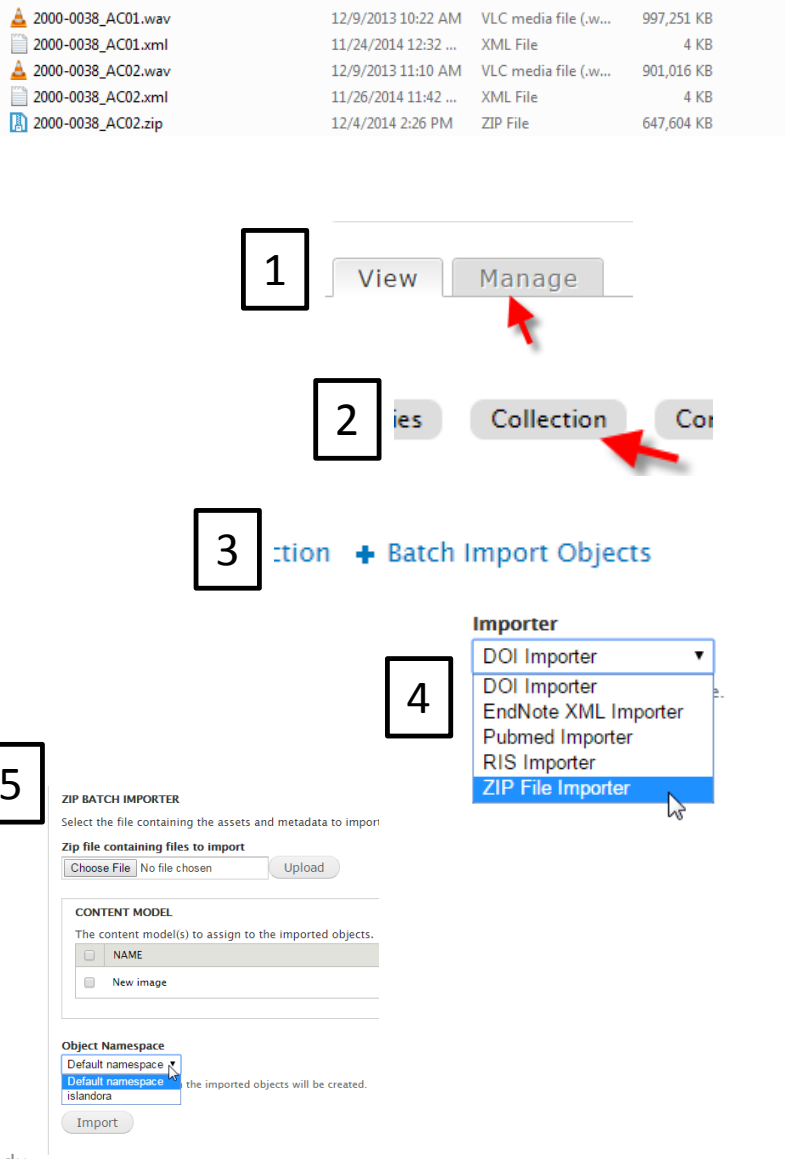
Add 1 Digital Asset

- Single upload
  - Click “Manage” (1)
  - Click “Add an object to this Collection” (2)
  - If you set your Collection for several Content models, you will need to select a Content Model to Ingest (i.e. you need to tell the system the format of your content, large image, basic image, PDF, audio, video). Once selected click “Next” (3)
  - You might have several custom Metadata forms. If so, then select a Metadata entry form (default Islandora form, CHI, or CTDA) (4)
  - Fill out by providing an appropriate description for your content and click “Next” by scrolling down to the bottom of the page (5)
  - Choose your file (6)
  - You can click “Upload” and then “Ingest” or just “Ingest”. There is no difference. (7)



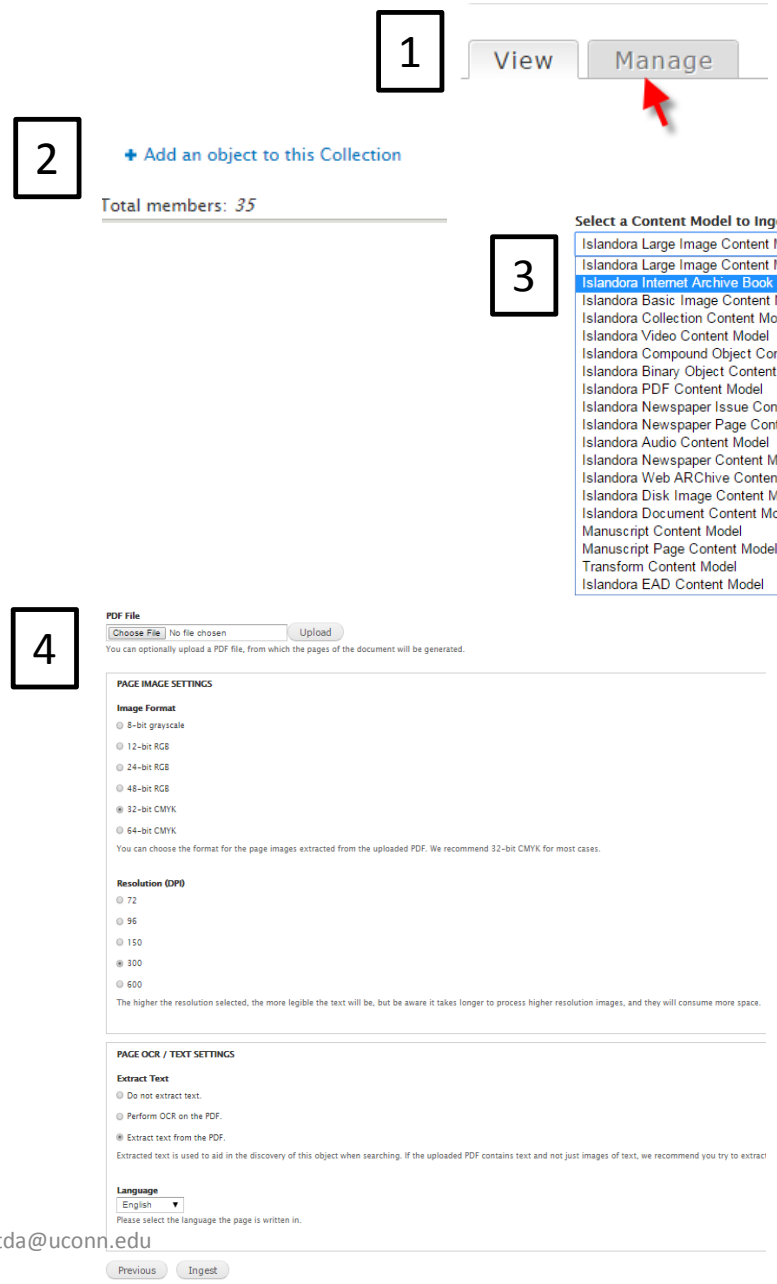
Batch Import Your Content

- Adding 1+ digital assets with or without associated metadata (Batch import)
  - All metadata files must be in XML and follow the CTDA MODS Implementation Guidelines (http://ctdigitalarchive.org/resources-for-participants/)
  - All XML files **MUST** be valid and well-formed
  - For content and their associated metadata files, the content file and MODS xml file **MUST** have the ***exact*** file name (i.e. myTiffimage01.tiff, myTiffimage.xml)
  - All files **MUST** be in a zipped folder (.zip)
- Go to where you want to add your content (which collection)
  - Click Manage (1)
  - Click Collection (2)
  - Select Batch Import Objects (3)
  - Select "ZIP File Importer" (4)
  - Choose your zip file (5)
  - Select the Content Model (all digital assets have to conform to the format of that Content Model) (5)
  - Select your namespace (5)
  - Click Import (5)



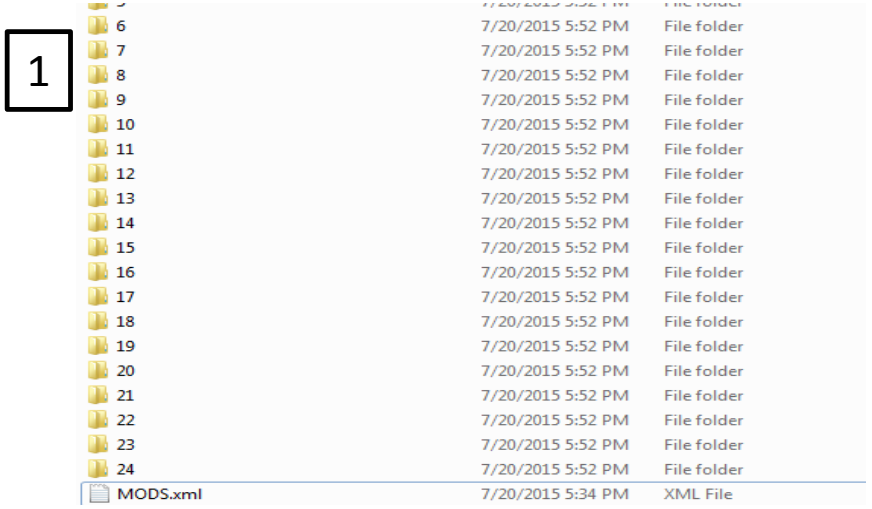
Adding A Book

- A Book is essentially a container that has pages. You can add pages in the form of a PDF or you can add pages in zip file. You can add one page at a time or multiple pages at different times.
  - Click Manage and then “Add an object to this collection” (1)(2)
  - Select “Islandora Internet Archive Book Content Model” and click Next (3)
  - Describe the book using the form and click Next
  - You can add the PDF of the book pages. If you do this, you can select the Page Image Settings, Resolution and Page OCR/Text Settings. Choose and upload your file. Select your settings. Click “Ingest” (4)
  - If you don’t have a PDF, then just select the language and click “Ingest” (4 bottom)



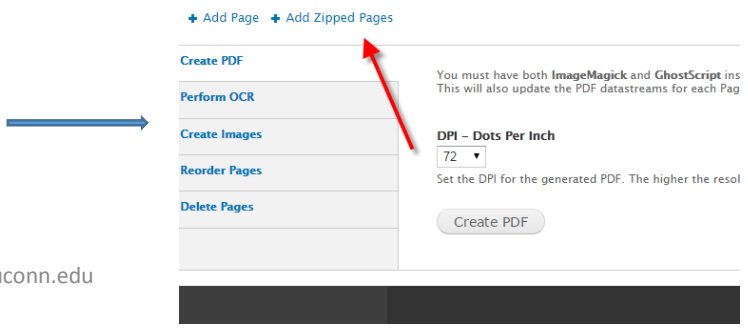
Adding A Book

- Book Settings
  - When you opt to add a PDF you can select image format, resolution and OCR
  - When you opt to load a zip file of pages (jp2 or tiff), you can select resolution only
  - Pages can be:
    - Loaded in different zip files – the sequence will follow the sequence of the image files in the zip file
    - Re-ordered
    - Deleted
  - OCR/HOCR is a datastream for the pages and not the Book
  - To have a downloadable PDF for the Book, you have to “Generate” this yourself
- Book batch: This works to ingest 1 book and its pages. There’s the MODS.xml file and each folder has the page number and in that folder is the page (1)



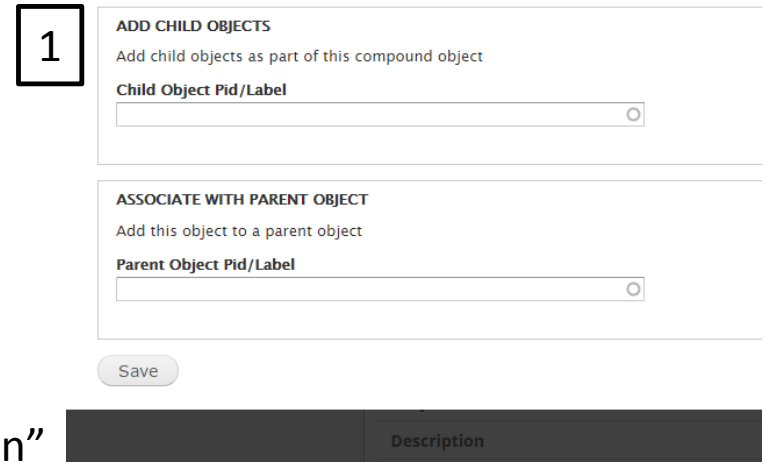
Adding A Newspaper

- Newspaper is a 3 step process. There is the newspaper, the issue and the pages of the issue.
  - Click Manage and “Add an object to this collection” and click Next
  - Select “Islandora Newspaper Content Model”
  - Describe the newspaper in the form and click “Ingest”. This will create a “container” for the issues of this newspaper
  - In this newspaper, click Manage and “Add Issue”. Describe the issue and click Next. You must add a W3C
  - You can add the PDF of the issue pages. If you do this, you can select the Page Image Settings, Resolution and Page OCR/Text Settings. Choose and upload your file. Select your settings. Click “Ingest”
  - If you don’t have a PDF, then just select the language and click “Ingest”
- Batch import pages of an issue: If you didn’t upload a PDF and have pages you want to batch import, go to the issue and click Manage. Go to the tab called “Issue” which opens a new window. Click “Add zipped pages”. Select your language and zip file to upload. Click upload and then “Add files”



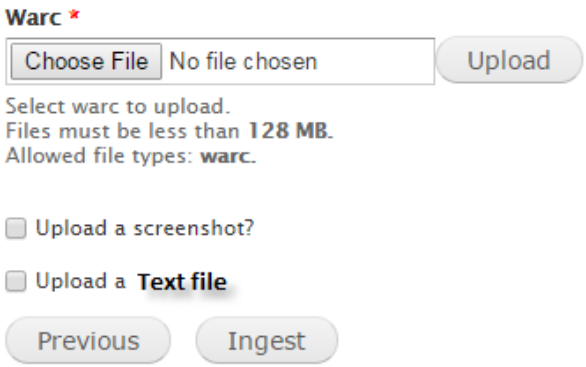
Adding A Compound Object

- Compound objects are parent/child relationships
- All of the child digital assets have to be uploaded before creating the compound object. You can put the child objects in a collection of their own if you want
- Note the PIDs of the digital assets that are to be children – you'll need to enter these PIDs later
- Create the parent.
  - Click Manage and "Add an object to this collection"
  - Select "Islandora Compound Object Content Model" and click Next
  - Describe the "parent" in the form and click "Ingest"
  - Click Manage again and click on the tab "Compound"
  - In the new window, enter the 1st PID of the child digital asset and click "Save". Repeat for all children. (1)



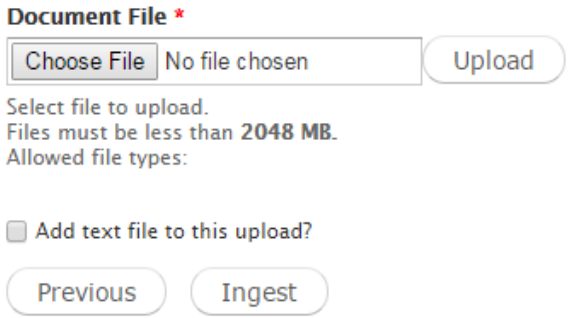
Adding A Web Archive File

- The Islandora Web Archive Content accepts warc file formats, a screen shot of that website and an associated text file.
- The text file needs to have the extension .txt.
- The Screen shot can be either an image that is the format jpg/jpeg or png.
- The warc format specifies a method for combining multiple digital resources into an aggregate archival file together with related information. It is a revision of the Internet Archives' ARC file format. More information can be found at: http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml
- If you have a warc file
  - Click Manage and “Add an object to this collection”
  - Select “Islandora WebARChive Content Model” and click Next
  - Describe the file and click “Next”
  - Select your warc file and click “Upload”
  - If you want to add a screen shot, click the box where you can select your image and upload. Same thing for your text file.



Adding A Document

- The Islandora Document Content Model is used for most documents such as word, text files, or pdf.
- Click Manage and “Add an object to this collection”
- Select Islandora Document Content Model and Next
- Describe the document and click “Next”
- Choose your
- If you have a PDF, odt, sxw, doc, docx, wpd, txt, html, ods, sxc, xls, csv, tsv, odp, sxi, ppt
  - Click Manage and “Add an object to this collection”
  - Select “Islandora Document Content Model” and click Next
  - Describe the file and click “Next”
  - Select your file and click “Upload”



Managing Datastreams

- There are several ways to manage information about your collection or digital asset
  - "Replace" the thumbnail for a collection
  - "Regenerate" a datastream
  - "Edit" the MODS (Metadata Object Description Schema) or descriptive metadata
  - "Delete" a datastream
  - Manage pages for books, newspapers, or manuscripts
  - Manage compound object relationships
  - Migrate content (called members) from one collection to another
  - Share content (called members) from one collection with another
  - Download datastreams
- If you "replace" or add a new MODS xml file, don't forget to copy the handle to paste that into your new MODS xml file. *The system will not reassign a handle – so don't forget to copy it over.*

+ Add a datastream

Bill Memorial Collection – 350002:BillMemorialCollection

ID	LABEL	TYPE	MIME TYPE	SIZE	VERSIONS	OPERATIONS			
RELS-EXT	Fedora Object to Object Relationship Metadata.	Inline XML	application/rdf+xml	564 B	1		download		
COLLECTION_POLICY	Collection policy	Inline XML	text/xml	2.36 KiB	3	replace	download		
MODS	MODS Record	Managed	application/xml	1.15 KiB	2	replace	download	edit	delete
DC	DC Record	Inline XML	text/xml	439 B	2	replace	download		
TN	Thumbnail	Managed	image/jpeg	255.57 KiB	3	replace	download		delete

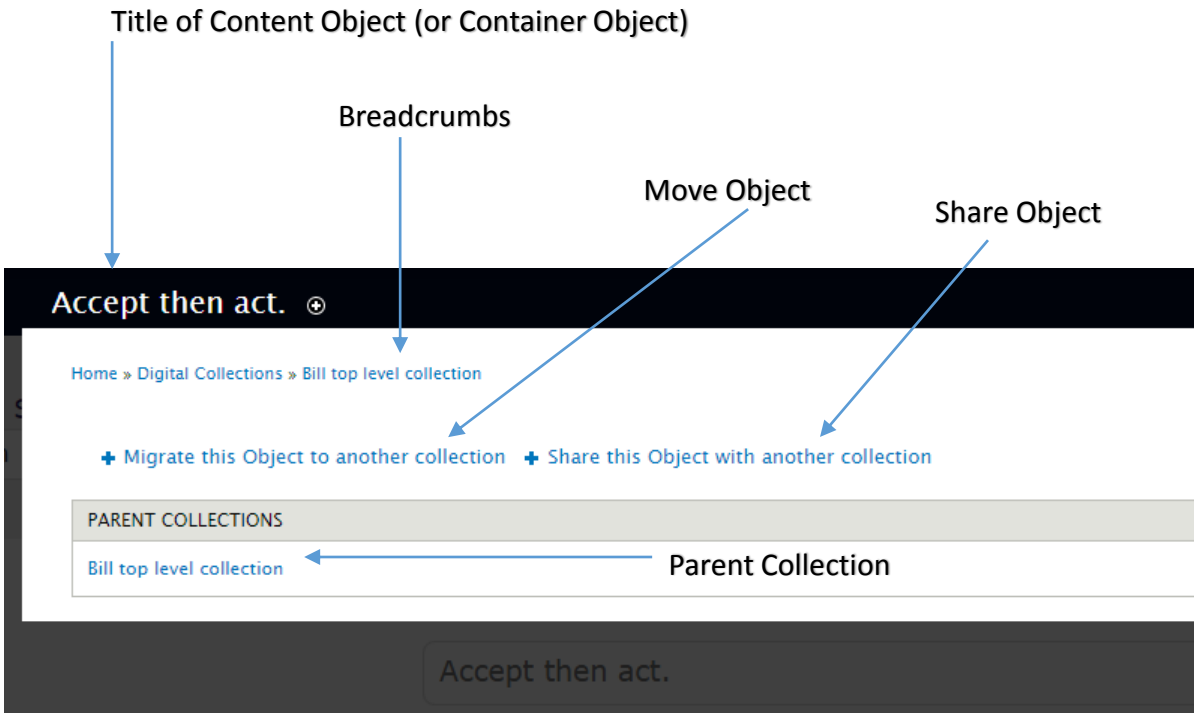
Managing Content in General

- There are several ways to manage information about your collection or digital asset
  - “Properties”: Edit the owner of an Object; Permanently remove an Object; Regenerate all derivatives.
  - “Collection”: Manage collection content model policies; Share members; Migrate members; Delete members; Batch import objects; Book batch (if collection is set for Internet Archive Book Content Model)
  - “Compound”: Manage compound relationships
  - “Book”: Add a page; Add a zip file of pages; Re-order pages; Create a PDF; Delete pages; Perform OCR; Update page thumbnails; Select page progression (right to left, left to right)
  - “Newspaper”: Add issue; Notification on any issues that don’t have a “YYYY-MM-DD” date
  - “Issue”: Add page; Add zipped pages; Create PDF; Perform OCR; Update page thumbnails; Page progression; Re-order pages; Delete pages
  - “Manuscript”: Create PDF; Perform OCR: Update page thumbnails; Page progression; Re-order pages; Delete pages

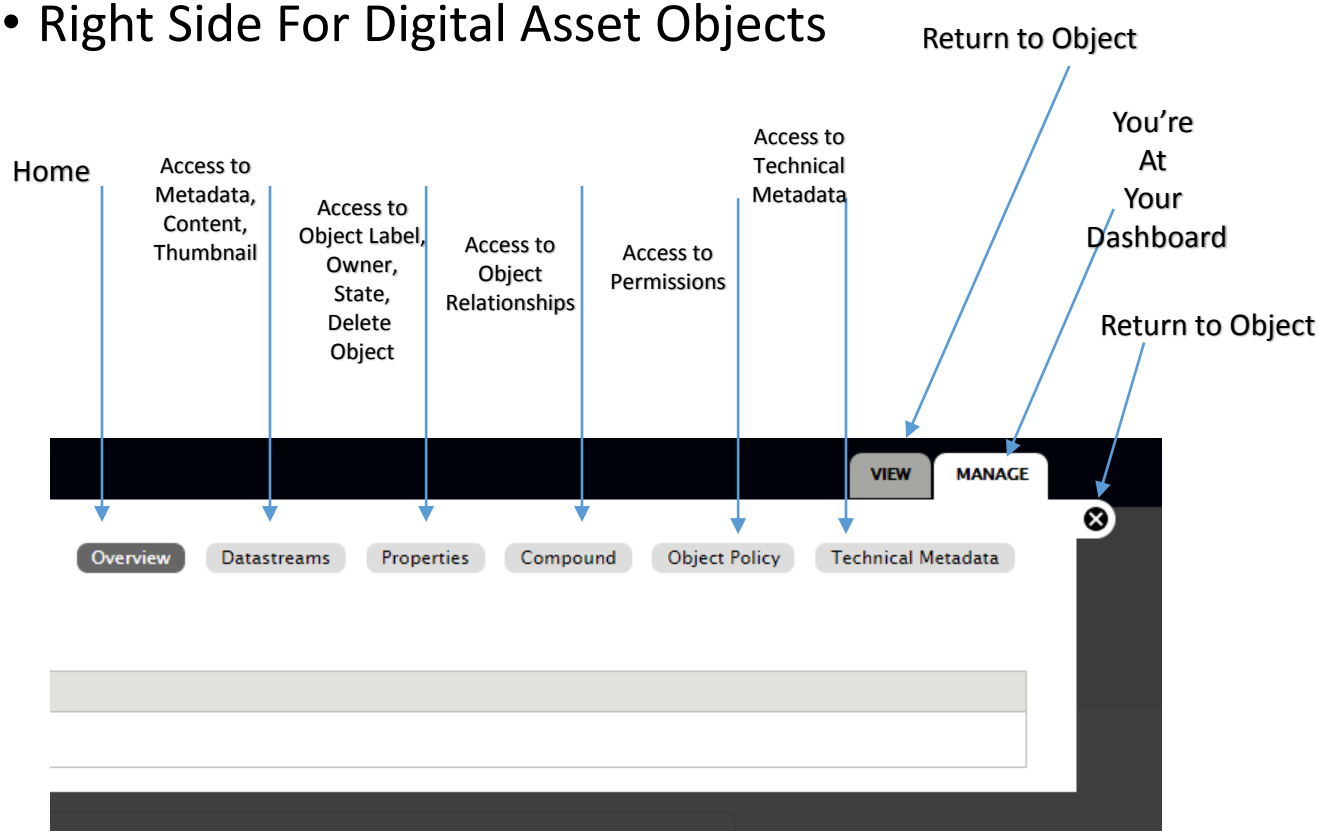
- Depending on the settings of your particular channel (site), there are also options for “Embargo”, “IP Embargo”, “Premis”, “Object Policy”, “TEI Transform”
- Not all options appear as many are dependent on the Content Model implemented. For example, if you added a tiff, Large Image Content Model, you will not see the “Book”, “Manuscript”, “Issue” or “Newspaper” options

Example of Manage Options

- Left side

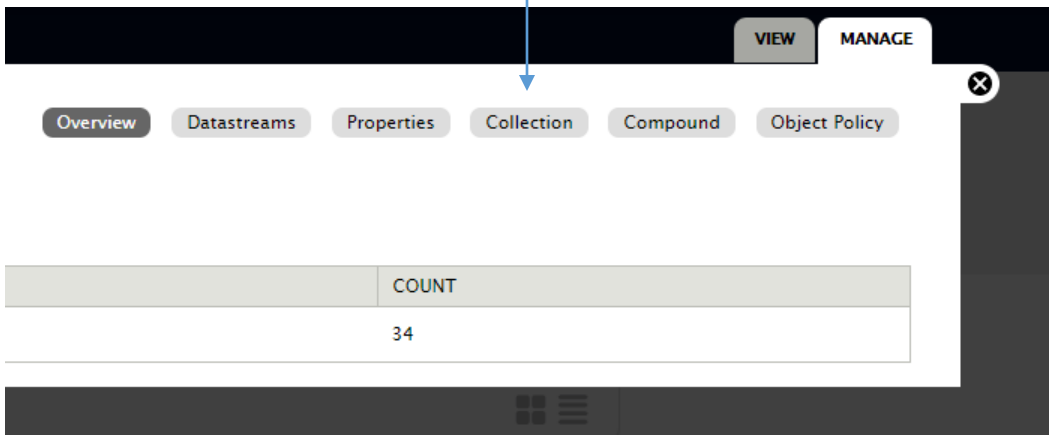


- Right Side For Digital Asset Objects

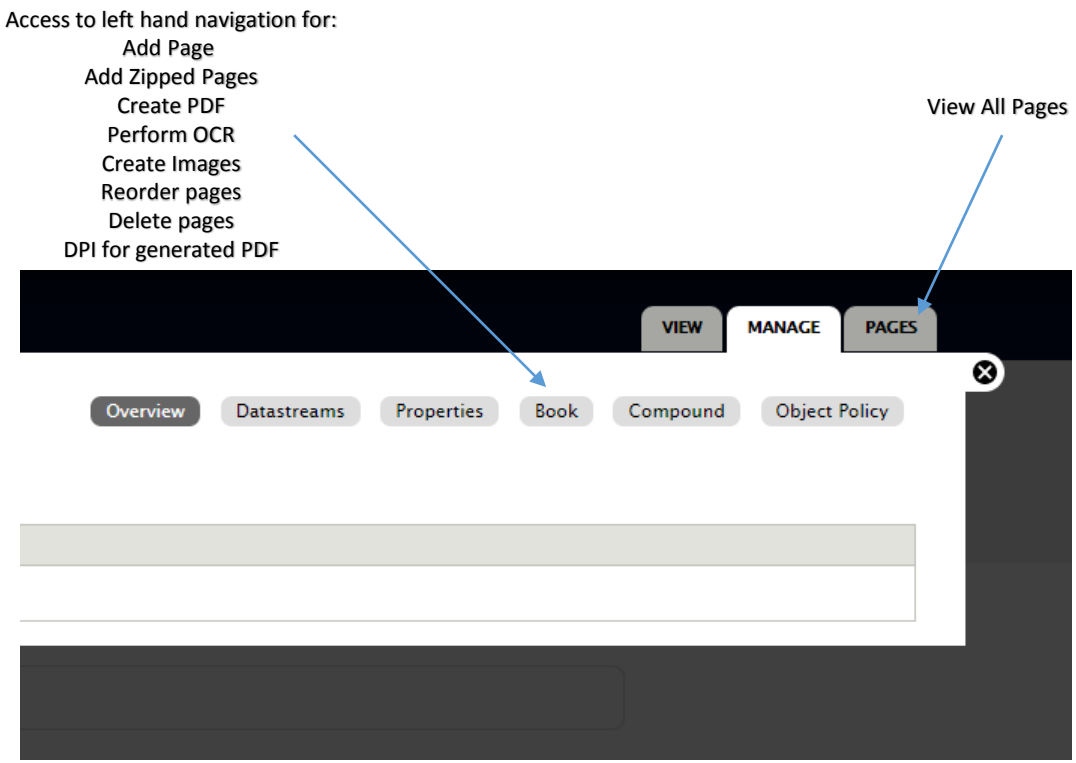


- Right Side for Collection Objects

Access to:
Collection Policies (i.e. Add or Remove Content Models)
Share Members
Migrate Members
Delete Members of Collection

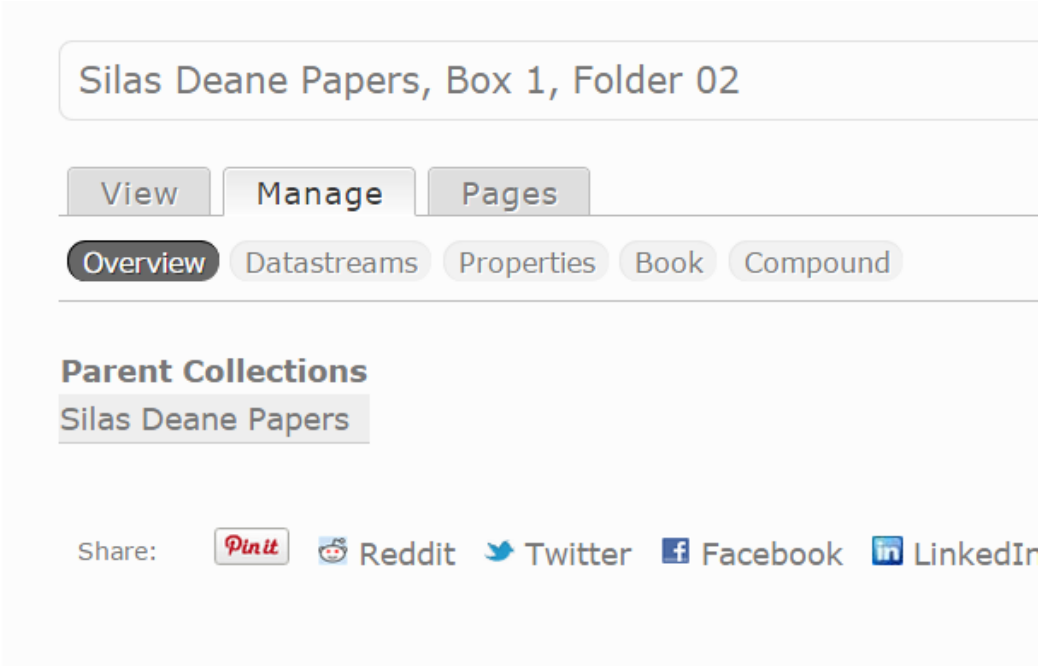


- Right side for book content models



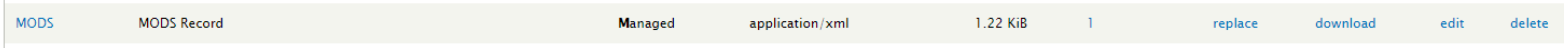
CAUTION

The Manage Options don't **ALWAYS APPEAR IN A POP UP WINDOW. THEY CAN APPEAR BELOW THE OBJECT'S TITLE AS SHOWN HERE**



CTDA Descriptive Metadata Standards

- CTDA uses the Metadata Object Description Schema (MODS) version 3.5 developed and maintained by the Library of Congress (http://www.loc.gov/standards/mods/)
- Cataloging Rules: Our implementation of MODS does not prescribe one set of cataloging rules such as Resource Description and Access (RDA) or Describing Archival Content Standard (DACS)
- Punctuation & HTML: Punctuation should only be retained if it occurs within an element and **SHOULD** be dropped between elements. Avoid the use of HTML.
- Single Upload will create a MODS datastream (record) that you can "Edit", "Replace", "Download" or "Delete"
- Batch Import and "Replace" require that the MODS datastream be an XML file



CTDA Minimal MODS Record

Required

- **Title**: Every record needs at least one title with no initial article for indexing purposes.
- **Resource Type**: Every record needs one of the following terms: text; cartographic; notated music; sound recording-musical; sound recording-nonmusical; sound recording; still image; moving image; three dimensional object; software, multimedia; mixed material. We use this information to facet searches.
- **Held By Statement:** Every record needs to indicate the name of the institution that acts as a steward for the digital resource.
- **Rights statement**: Every record needs to have a statement on use and reproduction.
- ***Persistent Identifier****: This is system generated.*
- **Language of MODS Record**: Your metadata form automatically adds this for you. If you create your own MODS XML records, you must include this information as ISO 639-2b.

Recommended

- **Digital Resource Origin**: Every record needs to indicate whether it was born digital or digitized.
- **Aggregation Tag**: There are 3 aggregation tags. One is to ensure content appears in Connecticut History Illustration channel. The other is to ensure content appears in a future map/geographic channel. The last is to ensure content appears in a future WWI channel.
- ***Date****:* To appear in the timeline as well as to ensure newspaper issues display correctly, dates should conform to the W3CDTF.
- ***Description****:* A brief abstract or description of the content (collection or digital asset) helps.
- ***Subjects****:* Most channels are set up for faceting searching based on topical and geographic (place) subjects.

Recommended Controlled Vocabularies & Data Standards

- Linked FAST (Faceted Application of Subject Terminology): Topics
- Library of Congress Linked Data Service: Authorities and Vocabularies: Names, Marc Relator Terms
- AAT (Art and Architecture Thesaurus): Genres
- TGN (Thesaurus of Geographic Names): Places
- Coordinates: Decimal format as Latitude, Longitude
- Topical Subjects: Select a term and avoid topical strings that are constructed such as Library of Congress Subject Headings that often mix topic and geographic terms
- Geographic Subjects: Select a place and avoid strings that are constructed such as Library of Congress Subject Headings that often mix topic and geographic terms

A Note on User Accounts

- Your Account is meant only for your use. If you need a separate account for guests, volunteers, or staff, contact ctda@uconn.edu
- If you are a site administrator, you must be aware of how CTDA manages users and user roles. This only applies to those who have their own channel.
- If you work in CTDA Manage, your institution can have up to 3 user accounts.

Resources

- Contact us with questions at: ctda@uconn.edu.
- http://ctdigitalarchive.org/resources-for-participants/
  - How to guides
  - MODS Implementation guides
  - Introduction to CTDA presentations
  - CTDigitalArchive Google Group
- CTDA Alert list: Please ask to be added to this list where all technical happenings that affect your work will be posted.