Fedora Web Client Batch Ingest v1.4

Below is a program overview followed by usage instructions.

The Batch Ingest process uploads Meta data files, and the data files they describe, into a Fedora Repository. All Meta data, is in the format of a METS record, and is normally contained in a single XML file. The files described by this META data are called Primary Content Objects (PCOs). A single METS file and all of its PCOs are ingested to become a single Fedora Digital Object (DO). Each DO is assigned a single unique Persistent IDentifier called a PID. In addition, a worldwide digital handle is also assigned by handle.net. The handle is requested from the Alliance Digital Repository (ADR) for each PID. This is accomplished by passing the PID to a web service hosted at the ADR http://adr.coalliance.org/ and the ADR obtains the handle from handle.net. The Ingest process consists of converting each METS file into a single FOXML file, which is then ingested, along with its PCOs into Fedora. Each PCO and the original METS file becomes a unique data stream within the Digital Object. The Batch Ingest program is packaged as a Web Application (WAR file) and deployed to a Tomcat web server. The applications defaults are stored in /WEB-INF/fedoraApp.properties within the WAR file.

The Batch Ingest is tied to the concept of a ***'batch set'***. A batch set is a single project for a particular institution. An example of an institution is the 'University of Denver' and a project would be a group of METs files and their PCOs. The institution and project (batch set) are mapped to a directory structure. If the parent folder, is '/batch\_space' and the institution is 'codu' and the project is 'frid', then the path to the batch set would be ***/batch\_space/codu/frid/*** . In this example ‘codu’ is the institution, Colorado University of Denver and ‘frid’ is the Fridlyand project.

Fedora Web Client Batch Ingest v1.47 may 2011

**NOTES:**

If ISLANDORA\_INGEST=true in the batchIngest.properites file, a collection and content model MUST be selected via form **OR** in the <mets:dmdSec ID="dmdAlliance"> section of a batch ingest file. This property also enable the collection and content model drop down lists on the batch ingest web interface.

Tasks is a new feature used to start an ingest without the use of the web interface. A timer is used to monitor the work directory for a batch-set. When a new batch file is detected, it is split and an ingest started. Indivual files can also be deposited in the mets/new or mets/updates for processing. This feature currently requires a unique class in the edu.du.penrose.systems.fedoraApp.tasks package with a coresponding property int the taskEnable.properties file. This feature will be used initially by the University of Denvers ECTD application. Once the bugs are worked out, the intention is to have a single generic class. The user will simply have to add a property that matches the batch-set name (see file structure below) to enable ingesting without the use of the web interface.

JMS code, to start tasks, rather then using a repeating timer., has been disabled due reliability issues.

The ability to Update has been added, however as of september 2011 this feature HAS NOT BEEN TESTED or USED IN PRODUCTION.

Updates as well and Adds are now displayed in the status screen.

**An example *batch set* directory (v1.47) and description follows…**

**batch\_space**

**codu (the institution,  *becomes fedora name space*)**

**frid (the project name, *becomes batch set name in ingest drop down menu*)**

**mets**

**new**

**updates**

**files**

**completed**

**completedBatchFiles**

**failed**

**failedBatchFiles**

**logs**

**work**

**images**

**tasksTemp**

**mets** - contains the METs XML files (any number) that are to be ingested.   
**files** - contains the PCOs (any number) referred to in the METs files.   
**completed** - METs files are moved into this directory after a successful ingest.   
**completedBatchFiles** – Batch files containing multiple mets are moved this directory after they have been split into individual files (stored in mets/new or mets/updates. NOTE: ***The ingest may still fail*** due to bad METS etc.  
**failed** - METs file are moved into this directory on an ingest failure.  
**failedBatchFiles** Batch files that created an error when we tried to split them are put here.  
**logs** - contains reports about each run of the batch ingest.   
**work** - used by the ingest process for temporary file storage. In addition a single XML file, which contains multiple METS records, can be stored here. After splitting the file (if it exists), the file is moved into the completed directory. If an error occurs during the split, the file is left in the work directory.  
**images used to add thumbnail to an ingest DEPRECIATED.  
tasksTemp used by tasks that monitory the files system looking for files to be ingested, rather being started an ingest manually from the web inteface. Tasks are enabled/disabled in the tasksEnable.properties file.**  
  
**Prior** to the start of a batch ingest, user options are performed. User options could include any of the following, deleting all previously completed files (in the completed folder), deleting all failed files (in the failed folder) and deleting previous reports (in the logs folder).   
  
**During** batch ingest, a METS file that ingests without error, is moved into the completed folder. If the user has selected to move PCOs, they will be moved into the completed folder along with its METS file. A METS file that fails is moved into the failed folder, the PCOs are not moved. *If an entire batch is aboted due to a FatalException all files are left in place!*

**After** completion of the batch ingest, two reports are generated, the main report which shows the number of completed and failed files, along with a list of all errors. The second report is a PID report. The PID report maps the PID and the contents of the OBJID element in the supplied METS file.

BATCH FILE SYNTAX

<batch version="2"> (1)

<batchDescription batchCreationDate="yyyy/MM/dd" > (0..1)

<literal>unformated text....</literal>

</batchDescription>

<ingestControl command="a | u" (add | update)

type for add is "normal | replyWithPid ; REQUIRED

type for updates is "all | meta | pco" ; REQUIRED

response="html status | pids to be assigned""> (1)

<mets> (0..\*)

..........

</mets>

</ingestControl>

</batch>

TYPICAL ISLANDORA BATCH FILE EXAMPLE

<batch version="2">

<ingestControl command="a " type=”normal”

<mets> (0..\*)

<mets:dmdSec ID="dmdAlliance">

….

<mets:mdWrap MIMETYPE="text/xml" MDTYPE="OTHER" LABEL="Custom Alliance Metadata">

<mets:xmlData>

<islandora collection="namespace:pid" contentModel="namespace:Whatever" />

</mets:xmlData>

</mets:mdWrap>

…..

</mets:dmdSec>

</mets>

</ingestControl>

</batch>

Batch Ingest Usage Instructions.

Use a web browser to connect to the Fedora Application. A typical URL will be <http://localhost/fedoraApp/> You will be presented with a Fedora login screen.om

|  |
| --- |
| ***Fedora Admin Client*** |
| |  | | --- | | [Home](http://localhost:8090/fedoraApp/hello.htm) | |
| |  |  | | --- | --- | | **Please Enter the Following Details** | | | **\*Protocol:** |  | | **\*Port:** |  | | **\*Server:** |  | | **\*User Name:** |  | | **\*Password:** |  | | **\* Means Required Field** | | |  | | |

Login to Fedora, using the appropriate settings for your setup.

Protocol http  
 Port 8080  
 Server localhost  
 Username fedoraAdmin  
 Password fedoraAdmin

After logging into fedora the following screen is displayed…

***Fedora Admin Client***

[**Logout**](http://localhost:8090/fedoraApp/logout.htm)[**Search Fedora objects**](http://localhost:8090/fedoraApp/searchFedora.htm) [**Get Fedora object**](http://localhost:8090/fedoraApp/getFedoraObj.htm) [**Fedora batch ingest**](http://localhost:8090/fedoraApp/batchIngest.htm)

Select the ‘Fedora batch ingest’ link.

The following screen is displayed…

***Fedora Admin Client***   
***1.2***

|  |
| --- |
| [Home](http://localhost:8090/fedoraApp/index.htm) |
| |  |  | | --- | --- | | **Perform Ingest of all files in the specified Batch Set.** | | | **\*Institution:** | | | **\*Batch set:** | | |  | **Stop on error during Ingest.** | |  | **Clear log files prior to Ingest.** | |  | **Clear previous failed files prior to Ingest.** | |  | **Clear previous completed files prior to Ingest.** | |  | **Moved Ingested PCOs to Completed Folder after Ingest.** | |  | **Split work/ to mets directory prior to ingest.** | | **Please select an Institution and Batch Set** | | |

First select the institution, followed by the institutions batch set. Once a valid batch set has been selected an [Ingest All] button will be displayed. Prior to starting the ingest make sure the correct options are set. The next page has a description of all *Batch Ingest* options.

VERSION 1.47 NOTES:

If ISLANDORA\_INGEST=true in the batchIngest.properites file, there will be additional collection and content model drop-down menus.

Updates as well and Adds are now displayed in the status screen.

* Stop on error during Ingest.

Normally the ingest will continue and all errors that occur will be displayed in the batch ingest report, however selecting this option will stop the batch ingest when the first error occurs.

* Clear log files prior to Ingest.

All log files (reports) from the previous run will be deleted prior to starting this ingest.

* Clear previous failed files prior to Ingest.

All failed files from the previous run will be deleted.

* Clear previous completed files prior to Ingest.

All files that successfully completed, from the previous ingest will be deleted.

* Move Ingested PCOs to Completed Folder after Ingest.

Upon successful ingest the PCOs will be moved to the completed folder along with their METS file.

* Split work/[insert file name] to mets directory prior to ingest.

This file will be split into multiple METS files and moved to the ingest directory (which may contain other single METS files) prior to ingest. NOTE: You must specify the file name AND check the checkbox.

The next page shows the *Batch Ingest* status screen.

***Fedora Admin Client***

|  |  |
| --- | --- |
| [Home](http://localhost:8090/fedoraApp/index.htm) | | [Start New Ingest](http://localhost:8090/fedoraApp/batchIngest.htm) |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | **File Count** | | | **Ingested:** | 0 | | **Failed:** | 0 | |   **Institution:   codu   Batch Set:   frid** |
| |  | | --- | | **Status:   Get PID and Handle for: may-15-2008-095159860.xml** | |

|  |
| --- |
|  |

If the [Stop Ingest] button is selected, the current ingest, of a METS record and it’s PCOs will be completed, and then the ingest halted.

After all files have been processed, the completion screen is displayed.

***Fedora Admin Client***

|  |  |
| --- | --- |
| [Home](http://localhost:8090/fedoraApp/index.htm) | | [Start New Ingest](http://localhost:8090/fedoraApp/batchIngest.htm) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **File Count** | | | **Ingested:** | 12 | | **Failed:** | 1 | |

|  |
| --- |
| **Institution:   codu   Batch Set:   frid** |
| |  | | --- | | **Status:   Completed** | |

|  |
| --- |
|  |
|  |
|  |

|  |
| --- |
|  |

The [Enable New Batch] button is used to enable a new run of this batch set. This is a precaution to avoid accidently restarting a batch set and removing the current reports, files etc.

**PID REPORT**

|  |  |  |
| --- | --- | --- |
|  |  |  |

Fedora PID, Mets OBJID, File Name

codu:716,frid00001,may-15-2008-095159860.xml

codu:717,frid00002,may-15-2008-095159875.xml

codu:718,frid00003,may-15-2008-095159891.xml

codu:719,frid00004,may-15-2008-095159907.xml

|  |  |  |
| --- | --- | --- |
| **INGEST REPORT** |  |  |

Start of Batch Ingest Report: Thu Jun 19 08:46:17 MDT 2008\*\*\*\*\*\*\*\*\*\*

Results -

Documents Completed = 12

Documents FAILED = 1

Failure Report -

Ingest FAILED for C:\batch\_space\codu\frid\mets\may-15-2008-095159922.xml Error: Unable to create FOXML DOM Error on line 29: The element type "mods:extentL" must be terminated by the matching end-tag "".

End of Batch Ingest Report ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

|  |
| --- |
|  |