



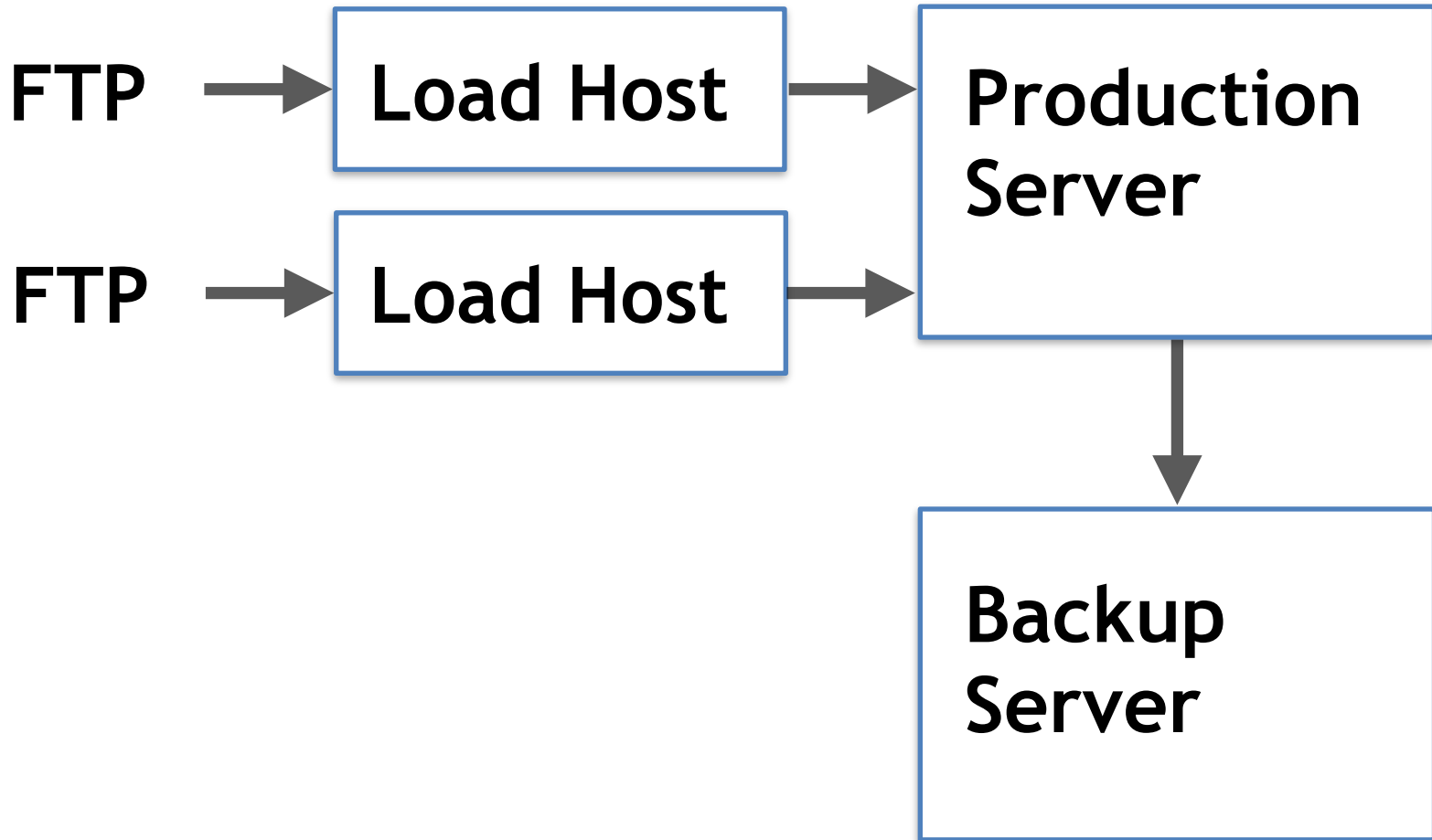
Experiences with Off-Production Server Ingests

**Randy Fischer, Programmer
Gainesville, Florida
May 6th, 2016**

Motivation

- The Florida Virtual Campus maintains an Islandora hosting service that currently serves 18 sites from a monolithic Islandora server.
- Ingests are ongoing, and because of the impact these ingests can have on the production services we have developed software to perform validation and derivative creation for packages on separate virtual hosts.

Architecture



FTP Home Layout

`/ftp/site/incoming/`

`/ftp/site/processing/`

`/ftp/site/warnings/`

`/ftp/site/errors/`

Package Handling - Watcher

god:
Watcher Process

.../incoming/PKG-ID/



*move to newly
created directory*

.../processing/aab/PKG-ID/



*enter directory
name in queue*

Redis Queue

Package Handling - Watcher

god config:

```
God.watch do |w|
  w.uid      = UID
  w.gid      = GID
  w.env      = { 'CONFIG_FILE' => "/usr/local/islandora/offline-ingest/config.yml" }
  w.name     = "ftp-handler"
  w.log_cmd  = "/bin/logger -t '#{w.name}' -p local1.info"
  w.start    = "/usr/local/islandora/offline-ingest/tools/ftp-handler"
  w.keepalive
end
```

Package Handling - Ingestor

god config:

```
God.watch do |w|
  w.uid      = UID
  w.gid      = GID
  w.env      = { 'CONFIG_FILE' => "/usr/local/islandora/offline-ingest/config.yml" }
  w.name     = "ingest-handler-1"
  w.group    = "ingest-handlers"
  w.log_cmd  = "/bin/logger -t '#{w.name}' -p local1.info"
  w.start    = "/usr/local/islandora/offline-ingest/tools/ingest-handler digitool ftp"
  w.keepalive
end
```

more stanzas

Package Handling - Ingestor

**god:
Ingest Process**

Redis Queue

*remove next directory
name from queue*

attempt ingest

*fail: move package
to /ftp/site/errors/*

*warn: move package
to /ftp/site/warnings/*

*succeed: delete
package*

write status info to a db

Package Handling - Status Report



write status info to db



Every affiliate site has a web site to report offline-ingest status

**A simple sinatra data browser service:
list, filter, drill down**

Package Handling - Status Report

Ingest Report for fsu.digital.flvc.org: All 34645 Digital Objects Selected

Date Status ID Title
Content Type Collection

<< First < Previous = Submit Clear Form = Next > Last >>

CSV: download these 34645 selected records

Date	Status	Package	Digitool	Title
2016-05-05 11:03:15	warning	FSU_WW2_16_0005 ⇒		Feldwebel Schaefer WWII 1940 Battle of France photograph album, 1941-1940
2016-05-05 07:56:12	warning	FSU_HPUA_catalog_1980_v73n5 ⇒		Florida State University Bulletin Catalog Issue
2016-05-05 08:49:57	warning	FSU_HPUA_catalog_1968_graduatebulletin ⇒		Florida State University Bulletin Graduate Issue
2016-05-04 21:30:37	warning	FSU_HPUA_catalog_lifeAtFSU ⇒		Life at FSU Bulletin
2016-05-04 20:05:11	warning	FSU_HPUA_catalog_1980_v74n3 ⇒		Florida State University Directory of Classes - Summer Quarter
2016-05-04 16:35:26	error	FSU_HPUA_catalog_1980_v73n5		Florida State University Bulletin Catalog Issue
2016-05-04 17:47:08	warning	FSU_HPUA_catalog_1980_v74n2 ⇒		Florida State University Schedule of Classes - Spring
2016-05-04 12:52:53	warning	FSU_HPUA_catalog_1979_v72n5 ⇒		Florida State University Bulletin Catalog Issue
2016-05-04 14:23:44	success	FSU_HPUA_catalog_1980_v73n2 ⇒		Florida State University Directory of Classes - Winter Quarter
2016-05-04 11:50:18	success	FSU_HPUA_catalog_1979_winterschedule ⇒		Florida State University Schedule of Classes - Winter Quarter
2016-05-04 11:50:42	warning	FSU_HPUA_catalog_1980_generalbulletin ⇒		Florida State University Bulletin General Information
2016-05-04 09:41:55	success	FSU_HPUA_catalog_1979_v73n1 ⇒		Florida State University Directory of Classes - Fall Quarter
2016-05-04 09:11:26	success	FSU_HPUA_catalog_1979_springschedule ⇒		Florida State University Directory of Classes - Spring Quarter
2016-05-04 09:33:03	error	FSU_HPUA_catalog_1979_v72n5		Florida State University Bulletin Catalog Issue
2016-05-03 17:58:47	warning	FSU_HPUA_catalog_1979_v72n4 ⇒		Florida State University Directory of Classes - Summer Quarter
2016-05-03 17:40:25	warning	FSU_HPUA_catalog_1978_winterschedule ⇒		Florida State University Schedule of Classes - Winter Quarter
2016-05-03 17:58:46	error	FSU_HPUA_catalog_1979_spring schedule		
2016-05-03 17:32:26	warning	FSU_HPUA_catalog_1978_summertentativelisting ⇒		Florida State University Tentative Course Listings - Summer Quarter
2016-05-03 16:11:14	warning	FSU_HPUA_catalog_1978_summerschedule ⇒		Florida State University Schedule of Classes - Summer Quarter
2016-05-03 15:31:54	warning	FSU_HPUA_catalog_1978_springschedule ⇒		Florida State University Schedule of Classes - Spring Quarter

Package Handling - Status Report

Ingest Failure Report for Package FSU_HPUA_catalog_1979_v72n5 on fsu.digital.flvc.org

[← Previous](#) = [Package List](#) = [Next →](#)

Title: Florida State University Bulletin Catalog Issue

Ingest Date: 2016-05-04 09:33:03

Processing Time: 0 seconds

Raw Ingest Size: 0.00 MB

Content Type: Book

Islandora Collections: [Florida State University Bulletins \(fsu:fsubulletins\)](#)

Package Errors

The Book Package FSU_HPUA_catalog_1979_v72n5 is missing the following 1 required file declared in the mets.xml file:

- FSU_HPUA_catalog_1979_v72n5_003%201.tif

Package Warnings

The Book Package FSU_HPUA_catalog_1979_v72n5 has the following 1 unexpected file that will not be processed:

- FSU_HPUA_catalog_1979_v72n5_003.tif

Package Handling - Status Report

Ingest Success Report for Package FSU_HPUA_catalog_1980_v73n2 on fsu.digital.flvc.org

← Previous = Package List = Next →

Title: Florida State University Directory of Classes - Winter Quarter

Islandora ID: [fsu:305678](#) ⇒



Ingest Date: 2016-05-04 14:23:44

Processing Time: 2 hours, 11 minutes, 40 seconds

Raw Ingest Size: 6334.07 MB

Content Type: Book

Component Objects (96): [fsu:305681](#), [fsu:305683](#), [fsu:305685](#), [fsu:305687](#), [fsu:305689](#), [fsu:305691](#), [fsu:305693](#), [fsu:305695](#), [fsu:305697](#), [fsu:305698](#), [fsu:305700](#), [fsu:305702](#), [fsu:305704](#), [fsu:305706](#), [fsu:305708](#), [fsu:305710](#), [fsu:305712](#), [fsu:305714](#), [fsu:305717](#), [fsu:305719](#), [fsu:305721](#), [fsu:305723](#), [fsu:305725](#), [fsu:305727](#), [fsu:305729](#), [fsu:305731](#), [fsu:305733](#), [fsu:305735](#), [fsu:305738](#), [fsu:305740](#), [fsu:305742](#), [fsu:305744](#), [fsu:305746](#), [fsu:305748](#), [fsu:305750](#), [fsu:305752](#), [fsu:305754](#), [fsu:305757](#), [fsu:305759](#), [fsu:305763](#), [fsu:305765](#), [fsu:305768](#), [fsu:305770](#), [fsu:305773](#), [fsu:305777](#), [fsu:305782](#), [fsu:305784](#), [fsu:305789](#), [fsu:305793](#), [fsu:305795](#), [fsu:305797](#), [fsu:305801](#), [fsu:305805](#), [fsu:305808](#), [fsu:305810](#), [fsu:305812](#), [fsu:305817](#), [fsu:305820](#), [fsu:305824](#), [fsu:305826](#), [fsu:305831](#), [fsu:305833](#), [fsu:305836](#), [fsu:305839](#), [fsu:305843](#), [fsu:305848](#), [fsu:305852](#), [fsu:305857](#), [fsu:305859](#), [fsu:305862](#), [fsu:305864](#), [fsu:305869](#), [fsu:305874](#), [fsu:305876](#), [fsu:305881](#), [fsu:305885](#), [fsu:305889](#), [fsu:305895](#), [fsu:305899](#), [fsu:305902](#), [fsu:305906](#), [fsu:305909](#), [fsu:305914](#), [fsu:305918](#), [fsu:305922](#), [fsu:305925](#), [fsu:305928](#), [fsu:305930](#), [fsu:305932](#), [fsu:305937](#), [fsu:305941](#), [fsu:305944](#), [fsu:305946](#), [fsu:305948](#), [fsu:305950](#), [fsu:305951](#)

Purl: http://purl.flvc.org/fsu/fd/FSU_HPUA_catalog_1980_v73n2

Islandora Collections: [Florida State University Bulletins \(fsu:fsbulletins\)](#)

Datastreams: [Archived METS for future reference](#), [Dublin Core Record](#), [MODS Record](#), [Relationships](#), [Table of Contents](#), [Thumbnail](#), [XACML Policy Stream](#)

Package Structure - Basic Image

what's this package thing anyway?

.../incoming/PKG-ID/

.../incoming/PKG-ID/PKG-ID.xml ***MODS***

.../incoming/PKG-ID/manifest.xml ***FLVC***

.../incoming/PKG-ID/image.foo ***Data...***

The manifest.xml file

```
<?xml version="1.0" encoding="UTF-8"?>
<manifest xmlns="info:flvc/manifest/v1">
  <contentModel>islandora:sp_pdf</contentModel>
  <collection>fau:bio</collection>
  <submittingInstitution>FAU</submittingInstitution>
  <owningUser>Randy Fischer</owningUser>
  <owningInstitution>FAU</owningInstitution>
  <objectHistory source="digitool">
    admin_unit="FAU01",
    ingest_id="ing11657",
    creator="creator:OCABEZAS",
    creation_date="2011-12-14 10:20:49",
    modified_by="super:FAUDIG",
    modification_date="2013-06-14 08:50:43"
  </objectHistory>
</manifest>
```

selected elements saved in MODS stream

The manifest.xml file

ELEMENT	REQUIRED?	REPEATABLE?
collection	yes	yes
contentModel	yes	
identifier		yes
label		
otherLogo		yes
owningInstitution	yes	
owningUser	yes	
submittingInstitution		
embargo		
pageProgression		

Packages: supported content models

- **basic image**
- **large image**
- **pdf**
- **books**
- **newspapers**

Packages: Books

.../incoming/PKG-ID/

.../incoming/PKG-ID/PKG-ID.xml

.../incoming/PKG-ID/manifest.xml

.../incoming/PKG-ID/mets.xml

.../incoming/PKG-ID/image-01

.../incoming/PKG-ID/image-02

.../incoming/PKG-ID/image-03

...USW

Code

github.com/FLVC/offline-ingest

- based on Hydra rubydora library (tuque in ruby)
- ingest daemon
- command line tools

Code

PackageClass (base class)



- **BasicImagePackage**
- **LargeImagePackage**
- **PDFPackage**

StructuredPageClass (base class)



- **BookPackage**
- **NewspaperPackage**

Advantages

- written in ruby
- handles failovers for image derivation (jp2)
- heavy lifting off main server (tesseract)
- reduces tedium of book, newspaper loading
- supports FLVC-specific data streams (e.g. TOC)
- load servers don't need drupal, fedora installs

Disadvantages

- written in ruby
- multiple datastream uploads (reindexing)
- all or nothing ingest
- errors are opaque to end users
- I have to maintain it



Questions?



**1753 W. Paul Dirac Drive
Tallahassee, Florida
www.flvc.org**