Automated Publishing of MorphoSource Audubon Cores via IPT

Presented by Warren Brown
January 23rd, 2020

MorphoSource RSS Feed

"MorphoSource provides summary reports of media, download usage, and download requests for media that represent specimens that have been reported to iDigBio. The media report is formatted according to the **Audubon Core metadata standard**, and so can be incorporated into publisher reporting software, such as an IPT. All report files are linked and described in a single RSS feed, which can be used to receive regular report updates via automated download. Additionally, all report files are individually listed below, sorted by iDigBio publisher and recordset. For each report, there is a link to: 1) a Comma Separated Values (CSV) spreadsheet with the primary media, download usage, or download request metadata; and 2) an XML file encoded in Ecological Markup Language (EML) providing metadata about the CSV spreadsheet. These reports are updated as necessary on a daily basis."

https://www.morphosource.org/About/report

FLORIDA MUSEUM

MorphoSource RSS Feed

Florida Museum of Natural History IPT Service (350857a8-4940-4d02-ad42-49557f8006f1)

Recordset	Media	Downloads	Download Requests	Pub Date
UF FLMNH Ichthyology (c38b867b-05f3-4733-802e-d8d2d3324f84)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:10:23 -0500
UF Florida Museum of Natural History Mammals (48e1b8c1-91aa-4b87-8ca0-de1f81232eaf)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:13:09 -0500
UF Florida Museum Paleobotany (0e162e0a-bf3e-4710-9357-44258ca12abb)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:30:32 -0500
University of Florida Herpetology (bd7cfd55-bf55-46fc-878d-e6e11f574ccd)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:17:02 -0500
University of Florida Invertebrate Paleontology (1ba0bbad-28a7-4c50-8992-a028f79d1dc5)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:27:15 -0500
University of Florida Vertebrate Paleontology (f9a33279-d6ba-41c7-a511-ef6adfcb6e20)	CSV EML	CSV EML	CSV EML	Sun, 19 Jan 2020 07:27:53 -0500

FLORIDA MUSEUM

MorphoSource RSS Feed

Thanks MorphoSource! We can leverage the RSS feed to complete a pipeline from MorphoSource to aggregators, such as iDigBio, via IPT!
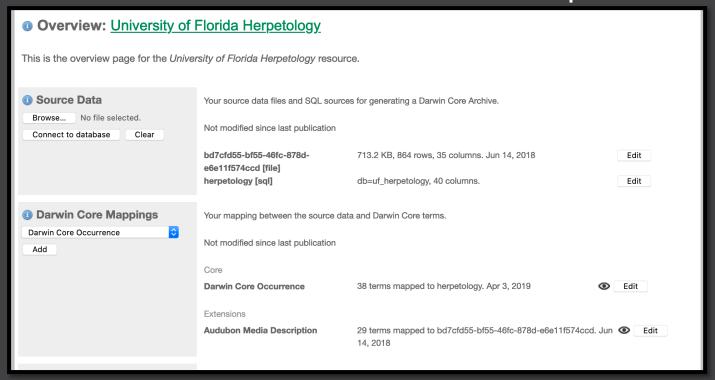
The methodology is straightforward:
1. Add an Audubon Media Description Extension to the Collection's IPT resource.
2. Use the MorphoSource RSS Downloader to automate update of the data source file on your IPT server.

Data changes in MorphoSource will feed into your data pipeline and reflect in aggregators without duplicate entry.
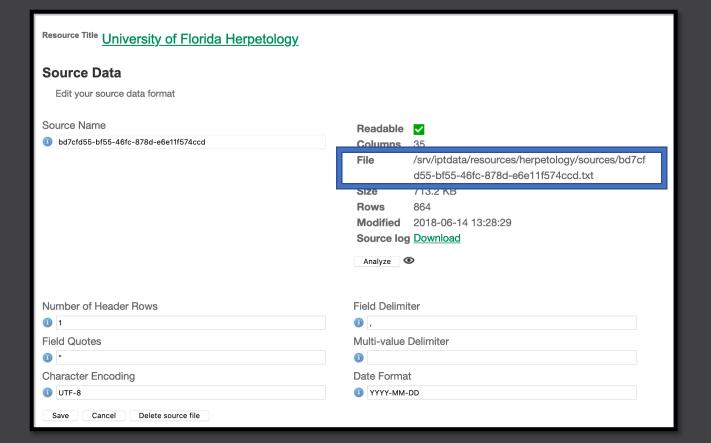
IPT Audubon Media Description

ⓘ Overview: University of Florida Herpetology

This is the overview page for the *University of Florida Herpetology* resource.

ⓘ Source Data

Browse... No file selected.

Connect to database   Clear

ⓘ Darwin Core Mappings

Darwin Core Occurrence

Add

Your source data files and SQL sources for generating a Darwin Core Archive.

Not modified since last publication

bd7cfd55-bf55-46fc-878d-e6e11f574ccd [file]   713.2 KB, 864 rows, 35 columns. Jun 14, 2018   Edit

herpetology [sql]   db=uf_herpetology, 40 columns.   Edit

Your mapping between the source data and Darwin Core terms.

Not modified since last publication

Core

Darwin Core Occurrence   38 terms mapped to herpetology. Apr 3, 2019   👁 Edit

Extensions

Audubon Media Description   29 terms mapped to bd7cfd55-bf55-46fc-878d-e6e11f574ccd. Jun 14, 2018   👁 Edit

Download the core file from MorphoSource and add it as a source data file in your IPT resource.
Then, add the Audubon Media Description extension and map the fields.

FLORIDA MUSEUM®

IPT Audubon Media Description

Resource Title University of Florida Herpetology

Source Data

Edit your source data format

Source Name

ⓘ bd7cfd55-bf55-46fc-878d-e6e11f574ccd

Readable ✅
Columns 35
File /srv/iptdata/resources/herpetology/sources/bd7cf
d55-bf55-46fc-878d-e6e11f574ccd.txt
Size 713.2 KB
Rows 864
Modified 2018-06-14 13:28:29
Source log Download

Analyze 👁

Take note of this path when adding the source data. You will need it later for the RSS Downloader.

Number of Header Rows

ⓘ 1

Field Quotes

ⓘ "

Character Encoding

ⓘ UTF-8

Field Delimiter

ⓘ ,

Multi-value Delimiter

ⓘ

Date Format

ⓘ YYYY-MM-DD

Save   Cancel   Delete source file

FLORIDA MUSEUM®

MorphoSource RSS Downloader

MorphoSourceRSSDownloader

MorphoSource RSS Downloader
Florida Museum of Natural History
Office of Museum Technology, 2018
Author: Warren Brown
https://floridamuseum.ufl.edu/omt

About

This application is intended to run as a scheduled task on an IPT server and downloads Audubon Core data files for a given collection. The application is a tool to automate the the flow of information as Morphosources releases new data.

The date of publication from the most recent download is recorded in a log file and the application only downloads a new file if the date of publication on the RSS feed is more recent than the recorded date. The intent is to schedule the application's execution to compliment the current Morphosource publication schedule.

https://github.com/FLMNH/MorphoSourceRSSDownloader

FLORIDA MUSEUM®

MorphoSource RSS Downloader

Usage

The application is a command line tool that accepts arguments of the a collection GUID, RSS feed address, and the path to which the downloaded file should be saved.

Information on the MorphoSource RSS feed is avaialable at https://www.morphosource.org/About/report and may be used to obtain your collection's GUID.

The syntax is

```
python morphosourcefeed.py COLLECTIONGUID FEEDADDRESS /PATH/TO/SAVE/FILE
```

PATH/TO/SAVE/FILE will be determined by your specific IPT server configuration and should be the "Source Data" path for your resource.

https://github.com/FLMNH/MorphoSourceRSSDownloader

FLORIDA MUSEUM®

MorphoSource RSS Downloader

Concrete Example

For a concrete example, this translates to the following for our IPT server:

```
cd /usr/local/sbin/MorphoSourceRSSDownload
source bin/activate
python morphosourcefeed.py "bd7cfd55-bf55-46fc-878d-e6e11f574ccd" "https://www.morphosource.org/rss/ms_rss
deactivate
```

YOUR PATH AND GUID WILL BE DIFFERENT.

https://github.com/FLMNH/MorphoSourceRSSDownloader

MorphoSource RSS Downloader

```
[wbrown@bacon cron.daily]$ cat MightyMorphoRSSRangers
#!/bin/bash

# UF Herps
cd /usr/local/sbin/MorphoSourceRSSDownloader && . bin/activate && python morphosourcefeed.py "bd7cfd55-bf55-46fc-8
78d-e6e11f574ccd" "https://www.morphosource.org/rss/ms_rss.xml" "/srv/iptdata/resources/herpetology/sources" && de
activate

# UF VP
cd /usr/local/sbin/MorphoSourceRSSDownloader && . bin/activate && python morphosourcefeed.py "f9a33279-d6ba-41c7-a
511-ef6adfcb6e20" "https://www.morphosource.org/rss/ms_rss.xml" "/srv/iptdata/resources/ufvp/sources" && deactivat
e

# UF Mammals
cd /usr/local/sbin/MorphoSourceRSSDownloader && . bin/activate && python morphosourcefeed.py "48e1b8c1-91aa-4b87-8
ca0-de1f81232eaf" "https://www.morphosource.org/rss/ms_rss.xml" "/srv/iptdata/resources/mammals/sources" && deacti
vate

# UF Paleobotany
cd /usr/local/sbin/MorphoSourceRSSDownloader && . bin/activate && python morphosourcefeed.py "0e162e0a-bf3e-4710-9
357-44258ca12abb" "https://www.morphosource.org/rss/ms_rss.xml" "/srv/iptdata/resources/ufpaleobotany/sources" &&
deactivate

#UF Fishes
cd /usr/local/sbin/MorphoSourceRSSDownloader && . bin/activate && python morphosourcefeed.py "c38b867b-05f3-4733-8
02e-d8d2d3324f84" "https://www.morphosource.org/rss/ms_rss.xml" "/srv/iptdata/resources/ichthyology/sources/" && d
eactivate

echo "Hello From MightyMorphoRSSDownloader"
[wbrown@bacon cron.daily]$
```

https://github.com/FLMNH/MorphoSourceRSSDownloader

FLORIDA
MUSEUM®

MorphoSource RSS Downloader

https://github.com/FLMNH/MorphoSourceRSSDownloader

Connect The Pipeline, It's Easy!

It is an easy lift to connect your MorphoSource data to your IPT server and bolster your collection's data pipeline!

Warren Brown
IT Director, Office of Museum Technology
Florida Museum of Natural History
wbrown@floridamuseum.ufl.edu