



**UNIVERSITY OF  
GEORGIA**

Chartered by the state of Georgia in 1785, the University of Georgia is the birthplace of public higher education in America — launching our nation’s great tradition of world-class public education. What began as a commitment to inspire the next generation grows stronger today through global research, hands-on learning and extensive outreach. A top value in public higher education, Georgia’s flagship university thrives in a community that combines a culture-rich college town with a strong economic center.

# Broken Links be Gone!

Creating a Broken Link Checker ArchivesSpace Plugin

*Corey Schmidt*  
*ArchivesSpace Project Manager*  
*University of Georgia Special Collections Libraries*  
[Corey.Schmidt@uga.edu](mailto:Corey.Schmidt@uga.edu)



UNIVERSITY OF  
GEORGIA



# Agenda

- **Background**
- **Why a plugin?**
- **Learning ArchivesSpace plugins**
- **Results**
- **Demo**
- **Resources**

# Background

- Note fields contain lots of links
- How to tell which links are broken?
- Web crawling with BeautifulSoup?
  - Too long to develop
  - Too long to run
- XML exports?
  - Long time to export from ArchivesSpace
  - Lots of local data
  - External script - not easy to run



# Why a Plugin?

- Pros:
  - Access the database directly, quickly
  - Part of ArchivesSpace, no external script/job
  - Easy for our staff to run anytime
- Cons:
  - No experience building a plugin
  - Worries about long-term maintenance

## Plugins and Scripts

 Created by Angela Spinazzé (Unlicensed)  
Last updated: Jan 26, 2022 by Christine Di Bella • min read • 208 people viewed

Thanks to everyone who has developed and shared a plugin or script. Have a plugin or script you'd like added to this list? Notice a mistake? Login or signup for an account you (christine.dibella@lyrasis.org) if you need help.

Homepage / Repository	Description
<a href="https://github.com/archivesspace/aspace-public-formats">https://github.com/archivesspace/aspace-public-formats</a>	ArchivesSpace plugin to expose raw metadata in the public interface
<a href="https://github.com/archivesspace/assets_reroute">https://github.com/archivesspace/assets_reroute</a>	A plugin to reroute the frontend and public UI asset pipeline ( needed for Tomcat7)
<a href="https://github.com/archivesspace/moma-ead-importer">https://github.com/archivesspace/moma-ead-importer</a>	Example of a customized EAD importer. Not intended to be used directly.
<a href="https://github.com/archivesspace/yale-ead-exporter">https://github.com/archivesspace/yale-ead-exporter</a>	Part of code4lib dev haus
<a href="https://github.com/uvalib-dcs/ASpace-plugins">https://github.com/uvalib-dcs/ASpace-plugins</a>	Some simple plugins for ArchivesSpace and notes about plugin layout.
<a href="https://github.com/hudmol/yale_containers">https://github.com/hudmol/yale_containers</a>	A new container model for ArchivesSpace
<a href="https://github.com/hudmol/aspace_yale_ils_integration">https://github.com/hudmol/aspace_yale_ils_integration</a>	An ArchivesSpace plugin that adds support for integrating Yale's ILS
<a href="https://github.com/hudmol/extended_advanced_search">https://github.com/hudmol/extended_advanced_search</a>	This is an ArchivesSpace plugin to extend the advanced search to index and allow searching on a series of record fields as specified by Yale University



# Learning ArchivesSpace Plugins

- Finding resources online
  - Set up developer environment for ArchivesSpace
  - Alicia Detelich's [custom report presentation](#)\*
  - Learning SQL, Ruby for more data parsing

\*Alicia Detelich, "Implementing a Custom Reports Plugin," ArchivesSpace Online Forum 2021, April 9, 2021, YouTube, 19:40. <https://www.youtube.com/watch?v=ruRWpOGaj1A>

```
175 def check_url(url, original_url, limit = 5)
176   begin
177     response_code = nil
178     if limit.zero?
179       uri = URI(original_url)
180       response = Net::HTTP.get_response(uri)
181       response_code = response.code.to_s
182     else
183       uri = URI(url)
184       response = Net::HTTP.get_response(uri)
185       case response
186       when Net::HTTPRedirection
187         location = response['location']
188         log("Following redirect #{location}")
189         check_url(location, original_url, limit - 1)
190       else
191         response_code = response.code.to_s if response.code.to_i != 200
192       end
193     end
194   rescue StandardError => e
195     response_code = e.message.to_s
196   ensure
197     response_code
198   end
199 end
200
```

# Results

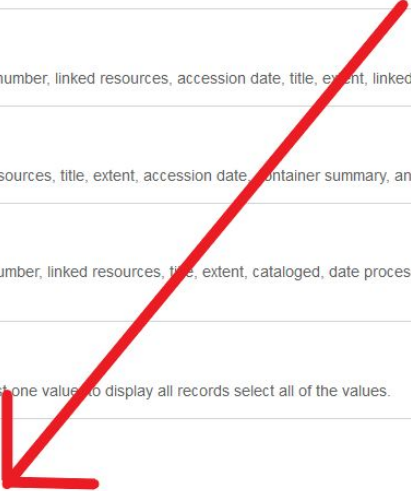
- Success! Installed plugin, tested, and deployed
- Reports on URLs and notes from:
  - Resources
  - Archival objects
  - Digital objects
  - Digital object components
  - Digital object file versions
  - Subject scope and contents
  - Agent person
  - Agent corporate entity
  - Agent family
  - Agent software
- Captures redirects up to 5 times
- Compatible with 2.8 and above
- Set as .CSV - use filtering in Excel

Failed to open TCP connection to find.sos.state.ga.us:80 (initialize: name or service not known)																				
certificate verify failed																				
404																				
Failed to open TCP connection to find.sos.state.ga.us:80 (initialize: name or service not known)																				
Failed to open TCP connection to find.sps.state.ga.us:80 (initialize: name or service not known)																				
404																				
Connection refused - Failed to open TCP connection to :80 (Connection refused - connect(2) for nil port 80)																				
503																				
Failed to open TCP connection to clinton.nara.gov:80 (initialize: name or service not known)																				
403																				
certificate verify failed																				
certificate verify failed																				
503																				
404																				
500																				
404																				
404																				
404																				
404																				
Connection refused - Failed to open TCP connection to :80 (Connection refused - connect(2) for nil port 80)																				
404																				
404																				
404																				
404																				
404																				
404																				
404																				
certificate verify failed																				
403																				
503																				
certificate verify failed																				

# Demo - 1

## New Background Job — Create Report Background Job

<b>Accession Receipt Report</b>	Select
Displays a receipt indicating accessioning of materials. Report contains accession number, title, extent, accession date, and repository.	
<b>Accession Report</b>	Select
Displays all accessions for a given repository. Report includes data entered into basic information fields for each accession.	
<b>Accessions Acquired and Linked Deaccession Records</b>	Select
Displays a list of accession record(s) and linked deaccession record(s). Report contains accession number, title, extent, accession date, container summary, cataloged, date processed, rights transferred, linked deaccessions and total extent of all deaccessions.	
<b>Accessions and Linked Subjects, Names and Classifications</b>	Select
Displays accessions and their linked names, subjects, and classifications. Report contains accession number, linked resources, accession date, title, extent, linked names, and linked subjects.	
<b>Accessions with Inventories</b>	Select
Displays only those accession records with an inventory. Report contains accession number, linked resources, title, extent, accession date, container summary, and inventory.	
<b>Accessions with rights transferred</b>	Select
Displays only those accession(s) for which rights have been transferred. Report contains accession number, linked resources, title, extent, cataloged, date processed, access restrictions, use restrictions, rights transferred and a count of the number of records selected with rights transferred.	
<b>Assessment Rating Report</b>	Select
Shows information for assessments with a specified rating in a specified range. You must select at least one value to display all records select all of the values.	
<b>Assessment Record List</b>	Select
All Created Assessment Records	
<b>Check URLs Report</b>	Select
Custom report to find URLs in notes, digital objects, and others and returns urls with error codes	





# Demo - 2

## New Background Job — Create Report Background Job

### Check URLs Report

Back

Custom report to find URLs in notes, digital objects, and others and returns urls with error codes

Format

CSV

Linked Records in JSON

Start Job

Cancel



# Demo - 3

Cancel Job  
**report\_job [131]** Background Job  
**Job Status**  

Status	<b>Running</b>
Time Submitted	2022-02-24 18:40:55 +0000
Time Started	2022-02-24 18:40:59 +0000

  
**Basic Information**  

Repository	Russell Library
Owner	cschmidt
Report Type	Check URLs Report
Format	CSV

Created by cschmidt 2022-02-24 18:40:55 +0000 | Last Modified by cschmidt 2022-02-24 18:40:55 +0000 | URI: /repositories/2/jobs/131

  
**Log** Follow Log

```
Generating report
Checking Digital Object File Versions...
Following redirect http://sclfind.libs.uga.edu/atfa/
Following redirect https://sclfind.libs.uga.edu/atfa/
Richard B. Russell Library for Political Research and Studies, election eve telecast '66, Digital Object File Version, http://podcaster.gcsu.edu/podcastdata/UGA/Channe
Richard B. Russell Library for Political Research and Studies, http://podcasting.gcsu.edu/4dcgi/podcasting/uga/episodes/28609/11700.mov, Digital Object File Version, h
Richard B. Russell Library for Political Research and Studies, http://podcaster.gcsu.edu/podcastdata/uga/channel_28099/podcast_8405/8405.mov, Digital Object File Versi
```



# Demo - 4

report\_job [131] Background Job

## Job Status

Status	<b>Completed</b>
Time Submitted	2022-02-24 18:40:55 +0000
Time Started	2022-02-24 18:40:59 +0000
Time Completed	2022-02-24 20:30:26 +0000

## Basic Information

Repository	Russell Library
Owner	cschmidt
Report Type	Check URLs Report
Format	CSV

Created by cschmidt 2022-02-24 18:40:55 +0000 | Last Modified by cschmidt 2022-02-24 18:40:55 +0000 | URI: /repositories/2/jobs/131

## Files

- [Download Report](#)

## Log

```
Following redirect https://bioguide.congress.gov/scripts/biodisplay.pl?index=B000940
Following redirect https://bioguide.congress.gov/search/bio/B000940
Done

Checking Agent Corporate Entity Notes...
Done

Checking Agent Family Notes...
No Repository, Breyfogle family, note_bioghst, findagrave.com, Connection refused - Failed to open TCP connection to :80 (Connection refused - connect(2) for nil po
Done

Checking Agent Software Notes...
Done

Adding report file.
```



# Demo - 5

350	Hargrett Library	ms1024	bioghist	<a href="http://www.newadvent.org">http://www.newadvent.org</a>	certificate verify failed
351	Hargrett Library	ms1109	bioghist	<a href="http://songwritershalloffame.org/index.php/exhibits/bio/">http://songwritershalloffame.org/index.php/exhibits/bio/</a>	404
352	Hargrett Library	ms1126	relatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1126/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1126/bmac-coll</a>	certificate verify failed
353	Hargrett Library	ms1174	separatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1174/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1174/bmac-coll</a>	certificate verify failed
354	Hargrett Library	ms1225	bioghist	<a href="https://doi.org/10.1093/ref:odnb/45508">https://doi.org/10.1093/ref:odnb/45508</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused
355	Hargrett Library	ms1243	bioghist	<a href="https://doi.org/10.1093/ref:odnb/4933">https://doi.org/10.1093/ref:odnb/4933</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused
356	Hargrett Library	ms1245	relatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1245/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1245/bmac-coll</a>	certificate verify failed
357	Hargrett Library	ms1283	separatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1283/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1283/bmac-coll</a>	certificate verify failed
358	Hargrett Library	ms1283a	scopecontent	<a href="http://www.lib.uga.edu/media/">http://www.lib.uga.edu/media/</a>	certificate verify failed
359	Hargrett Library	ms1310	bioghist	<a href="http://www.cviog.uga.edu/about/chapel/history.php">http://www.cviog.uga.edu/about/chapel/history.php</a>	404
360	Hargrett Library	ms1343	bioghist	<a href="http://www.csrees.usda.gov/qlinks/extension.html">http://www.csrees.usda.gov/qlinks/extension.html</a>	Failed to open TCP connection to www.csrees.usda.gov:80 (initialize: name or s
361	Hargrett Library	ms1343	bioghist	<a href="http://www.whitehouse.gov/about/presidents/woodrow">http://www.whitehouse.gov/about/presidents/woodrow</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused
362	Hargrett Library	ms1344	relatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1344/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1344/bmac-coll</a>	certificate verify failed
363	Hargrett Library	ms1371	bioghist	<a href="http://www.uga.edu">http://www.uga.edu</a>	certificate verify failed
364	Hargrett Library	ms1385	bioghist	<a href="http://www.encyclopediavirginia.org/Cooke_John_Esten_">http://www.encyclopediavirginia.org/Cooke_John_Esten_</a>	certificate verify failed
365	Hargrett Library	ms1406	bioghist	<a href="http://library.oglethorpe.edu">http://library.oglethorpe.edu</a>	certificate verify failed
366	Hargrett Library	ms1465	bioghist	<a href="https://tshaonline.org/handbook/online/articles/fca02">https://tshaonline.org/handbook/online/articles/fca02</a>	certificate verify failed
367	Hargrett Library	ms1478	bioghist	<a href="http://mgagnon.myweb.uga.edu">http://mgagnon.myweb.uga.edu</a>	Failed to open TCP connection to mgagnon.myweb.uga.edu:80 (initialize: name
368	Hargrett Library	ms1786	odd	<a href="https://galileo-usg-uga-primo.hosted.exlibrisgroup.com/p">https://galileo-usg-uga-primo.hosted.exlibrisgroup.com/p</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused
369	Hargrett Library	ms1988	relatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms1988/bmac-coll">http://purl.lib.uga.edu/brown/har-ms1988/bmac-coll</a>	certificate verify failed
370	Hargrett Library	ms2003	relatedmaterial	<a href="http://purl.lib.uga.edu/brown/har-ms2003/bmac-coll">http://purl.lib.uga.edu/brown/har-ms2003/bmac-coll</a>	certificate verify failed
371	Hargrett Library	ms2066	bioghist	<a href="http://www.georgiaencyclopedia.org/file/11751">http://www.georgiaencyclopedia.org/file/11751</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused
372	Hargrett Library	ms2106	bioghist	<a href="http://www.athensclarkecounty.com/history/unify.htm">http://www.athensclarkecounty.com/history/unify.htm</a>	Connection refused - Failed to open TCP connection to :80 (Connection refused

# Resources

- “ArchivesSpace Plug-ins.” ArchivesSpace tech-docs.  
<https://archivesspace.github.io/tech-docs/customization/plugins.html>
- Clair, Kevin. “Writing ArchivesSpace plugins: a reasonably comprehensive introduction.” University of Denver ArchivesSpace User Group. August 2, 2016. <https://slideplayer.com/slide/11648048/>
- Detelich, Alica. *Implementing a Custom Reports Plugin*. ArchivesSpace Online Forum 2021. April 9, 2021. YouTube, 19:40. <https://www.youtube.com/watch?v=ruRWpOGaj1A>
- “Hello World ArchivesSpace Plug-in.” GitHub repository. ArchivesSpace. Last modified May 1, 2018.  
[https://github.com/hudmol/archivesspace/tree/master/plugins/hello\\_world](https://github.com/hudmol/archivesspace/tree/master/plugins/hello_world)
- “Resources for developers.” ArchivesSpace tech-docs.  
[https://archivesspace.github.io/tech-docs/readme\\_develop.html](https://archivesspace.github.io/tech-docs/readme_develop.html)
- “Ruby Tutorial.” Tutorialspoint. 2021. <https://www.tutorialspoint.com/ruby/index.htm>
- University of Georgia Libraries. “uga-archivesspace-reports.” GitHub repository. University of Georgia. Last modified January 24, 2022. <https://github.com/uga-libraries/uga-archivesspace-reports>



# *Thank You!*

*Corey Schmidt  
ArchivesSpace Project Manager  
University of Georgia Special Collections Libraries  
[Corey.Schmidt@uga.edu](mailto:Corey.Schmidt@uga.edu)*

