Batch Processes for Faculty Work

Gilbert Borrego, Christopher Kehoe,
Colleen Lyon, & David Roberts
University of Texas at Austin

Texas ScholarWorks

- Created in 2008

- Consists mostly of library managed collections

- Content from all areas of the University

- Not a lot of faculty content

Texas ScholarWorks

- Conference proceedings/journals: 159

- Student Work (not ETDs): 1525

- UT Communities: 11949

- ETDs: 17983

- Faculty/Researcher Works: 2609

- Total: 34225

Faculty publications

Hard to demonstrate our value without use

We Decided to Proactively Add Content

- Faculty/researcher published articles

- Where allowed by license (CC-BY) or by publisher policy

- Only PDFs for now

- Created an FAQ to handle any questions about content

Where do we find this content & licensing info?

Using Google Script to Automate Sherpa/Romeo Copyright Check

SHERPA/RoMEO is a service run by SHERPA to show the copyright and open access self archiving policies of academic journals.

SHERPA/RoMEO

Journal: **Astrophysics** (ISSN: 0571-7256, ESSN: 1573-8191) [A translation of the Russian language j
RoMEO: This is a RoMEO **green** journal
Paid OA: A paid open access option is **available** for this journal.

Author's Pre-print: ✔ author **can** archive pre-print (ie pre-refereeing)
Author's Post-print: ✔ author **can** archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF: ✘ author **cannot** archive publisher's version/PDF

General Conditions:
- Author's pre-print on pre-print servers such as arXiv.org
- Author's post-print on author's personal website immediately
- Author's post-print on any open access repository after 12 months after publication
- Publisher's version/PDF cannot be used
- Published source must be acknowledged
- Must link to publisher version
- Set phrase to accompany link to published version (see policy)
- Articles in some journals can be made Open Access on payment of additional charge

Mandated OA: (Awaiting information)
Paid Open Access: Open Choice
Copyright: Self-archiving policy - Authors Rights - Funder Compliance
Updated: 16-May-2014 - Suggest an update for this record
Link to this page: http://www.sherpa.ac.uk/romeo/issn/0571-7256/

Published by: Springer Verlag (Germany) - Green Policies in RoMEO

Using Google Script to Automate Sherpa/Romeo Copyright Check

- Find affiliation on [Web of Science](#)

- Problem: WoS only allows a download of 500 records at a time.

- Import Web of Science .csv list to Google Spreadsheet

Using Google Script to Automate Sherpa/Romeo Copyright Check

- Insert Google Script created by Stephen X. Flynn (The College of Wooster) into Google script editor

- Register for an API key then insert the key below in the "Sherpa API key" section of the code. Without an API key, you will be limited to 500 requests per day, per IP address.

- Add and check the four functions produced (pubpdf, finaldraft, embargo, checkOAmandate)

- Create or sort list according to pubpdf for lowest hanging fruit

	ISSN	pdf	finaldraft	embargo
Cell	10591524	Publisher's version/PDF may be used	NO final draft allowed	no
	17520894	No publisher PDF	Final draft restricted	Embarg
	10736700	No publisher PDF	Final draft restricted	Embarg
	10926690	Publisher's version/PDF may be used	Final draft allowed	no
	3051048	Publisher's version/PDF may be used	Final draft allowed	no
ons and Me	1432087	No publisher PDF	Final draft restricted	Embarg
ent	10860266	No publisher PDF	Final draft allowed	Embarg
bclimatolog	310182	No publisher PDF	Final draft allowed	no
	310239	No publisher PDF	Final draft restricted	Embarg
	8831351	Unknown	Final draft restricted	no
	314528	Unknown	Final draft allowed	no
sychology B	1461672	No publisher PDF	Final draft allowed	Embarg
	15191397	Unknown	status unknown	no

Processing
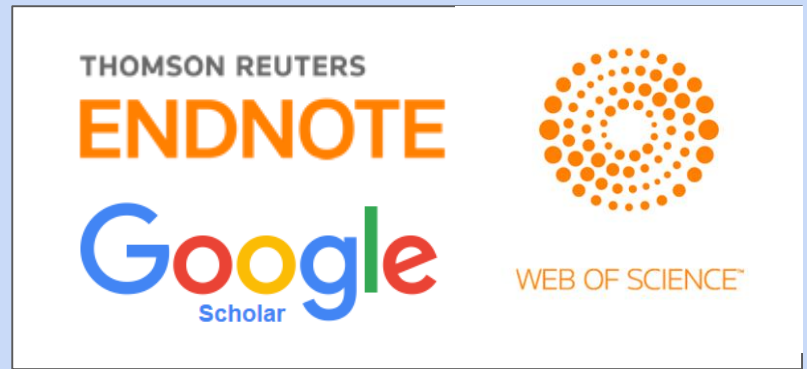
- Double check copyright & permissions

- Downloading metadata from Web of Science and open with EndNote

- Export EndNote data to Excel

- Find and download articles freely available from WoS or Google Scholar

- Input missing metadata into and 'clean-up' spreadsheet prior to batch upload

Finished Batch Ready for Upload

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	dc.identifier.Fil	dc.title	dc.creator	dc.type	dc.title.serial	dc.iden	created	dc.righ	dc.lang	dc.iden	dc.description.	dc.identifier.citatic	dc.description.sponsc	dc.contributtion.	nts.rest	dc.subject		dc.description.abstract
185	2005_Hoffmeyer.p	In Vitro Analysi	Hoffmeyer, Micha	Article	Cancer Cell Inte	1475-286	2005	Adminis	English	10.1186/1	Cellular and Mole	Hoffmeyer, Michaela F	NIH/NCI CA83957-01A1 Ur	Hoffmeyer, Michaela	Open	oncology		Background: Inflammatory br
186	2015_05_Aiken.pd	Triggered Seis	Aiken, Chastity||Zi	Article	Bulletin of the S	0037-110	2015-05	Adminis	English	10.1785/0	Geological Scienc	Aiken, Chastity, Jessic	National Science Foundatic	Walter, Jacob I.	Open	non-volcanic tremor||9.0 tohoku		We conduct a systematic sea
187	2015_09_Allshous	Lagrangian Ba	Allshouse, Michae	Article	Chaos	1054-150	2015-09	Adminis	English	10.1063/1	Nonlinear Dynami	Allshouse, Michael R.,	ONR N000141210665	Allshouse, Michael F	Open	gulf-of-mexico||fuzzy c-means||f		There has been a proliferation
188	2015_08_Allshous	Refining Finite	Allshouse, Michae	Article	Chaos	1054-150	2015-08	Adminis	English	10.1063/1	Nonlinear Dynami	Allshouse, Michael R.,	ONR N000141210665	Allshouse, Michael F	Open	lagrangian coherent structures||		While more rigorous and soph
189	2007_12_Askan.pc	Full Waveform	Askan, Aysegul||A	Article	Bulletin of the S	0037-110	2007-12	Adminis	English	10.1785/0	Geological Scienc	Askan, Aysegul, Volka	NSF ITR 0326449, EAR 01C	Ghattas, Omar	Open	travel-time tomography||earth st		We present a least-squares o
190	1994_09_Barkley.p	A Dynamical S	Barkley, Dwight||K	Article	Chaos	1054-150	1994-09	Adminis	English	10.1063/1	Nonlinear Dynami	Barkley, Dwight, and lo	NSF DMS 92-06224, ECS 9	Barkley, Dwight	Open	mathematics, applied||physics, r		A simple system of five nonlir
191	2012_12_Blazevski	Using Scatterir	Blazevski, Daniel||	Article	Chaos	1054-150	2012-12	Adminis	English	10.1063/1	Mathematics	Blazevski, Daniel, and	NSF DMS0901389||Texas (	Blazevski, Daniel||Fra	Open	phase-space||mathematics, app		Scattering theory is a conveni
192	2012_03_Brager.pc	Impaired Dendi	Brager, Darrin H.||/	Article	Cell Reports	2211-1247	2012-03	Adminis	English	10.1016/j.	Learning and Merr	Brager, Darrin H., Arvir	FRAXA||University of Texa	Brager, Darrin H.||Akt	Open	mental-retardation protein||ca1 p		Despite extensive research in
193	2013_08_Cai.pdf	Functional Act	Cai, Bowen W.||Ju,	Article	Cell Reports	2211-1247	2013-08	Adminis	English	10.1016/j.	Cellular and Mole	Bowen, Cai, Jeong-Ho	CA132813, NCI NIH HHS||C	Lee, Jeong-Ho||Paull	Open	base excision-repair||double-str		The prostate tumor suppress
194	2015_09_Clark.pdf	Evidence for P	Clark, Kate J.||Hay	Article	Bulletin of the S	0037-110	2015-09	Adminis	English	10.1785/0	Geophysics, Institr	Clark, Kate J., Bruce W	GNS||EQC Biennial Project	Wallace, Laura M.	Open	north-island||hawkes bay||great (		At the southern Hikurangi ma
195	2015_09_Corwin.p	Modeling Cour	Corwin, Lisa A.||Gr	Article	Cbe-Life Scienc	1931-7913	2015-03	Adminis	English	10.1187/c	Science and Math	Corwin, Lisa A., Mark J	Howard Hughes Medical In	Corwin, Lisa A.||Dolar	Open	genomics education partnership		Course-based undergraduate
196	2015_12_Corwin.pc	The Laborator	Corwin, Lisa A.||Rr	Article	Cbe-Life Scienc	1931-7913	2015-12	Adminis	English	10.1187/c	Educational Psych	Corwin, Lisa A., Christc	National Science Foundatic	Corwin, Lisa A.||Runy	Open	undergraduate research experier		Course-based undergraduate

Issues and Things to Consider

- Web of Science changes frequently

  - Finicky, the kind of metadata it provides can vary

- Time spent manually inputting some metadata

  - All caps, accent marks, **||**

  - Finding and adding citations

  - Sources of funding in WoS

- Journal and publisher copyrights change often

  - Good faith - Check SHERPA/RoMEO often

Getting content into DSpace

- Until recently - used a program created in-house to generate simple archive format (SAF)

- Repository staff run program and hand off SAF to Texas Digital Library (TDL) for ingest into repository

- TDL still had to test batches and write scripts to get content into DSpace

Problems with Process

- Batch process running slow

    - Better with SAF Creator - much faster

- Upgrade halted process of batches

- Still potential for negative feedback from faculty
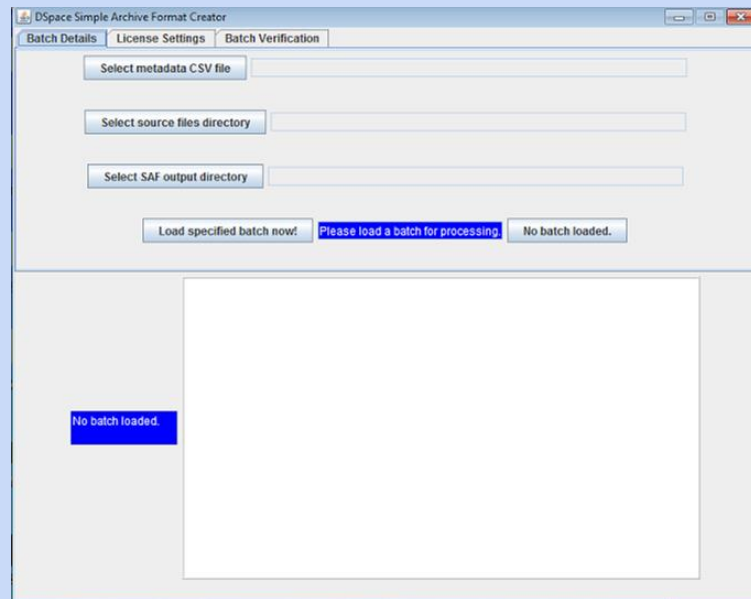
Results

- Currently 2609 items in Faculty/Research collection

    - Of these, 1695 have been added in the last 18 months

- 718 waiting to be ingested

- 262,647 downloads since April 2014

- 0 negative comments

Next Steps

- Continue processing batches

  - Go beyond just the PDFs

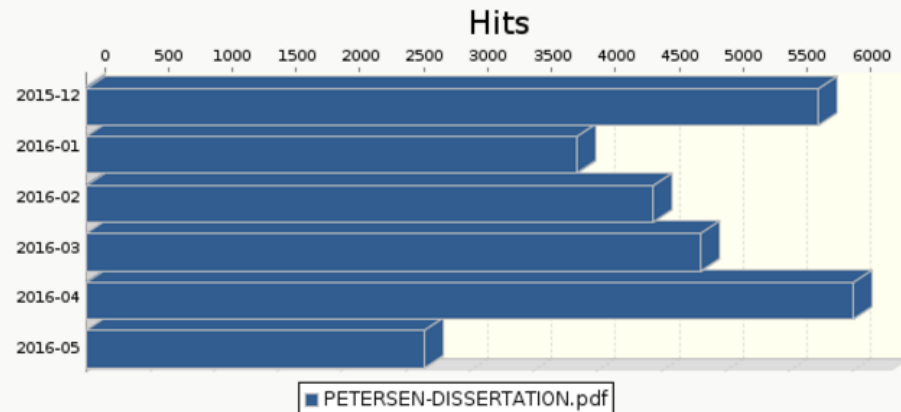- Use usage statistics to recruit more content

- Hire more students

Last month's file downloads

	2016-04
PETERSEN-DISSERTATION.pdf	6,014

Top 5 File downloads for the last 6 months

Hits

PETERSEN-DISSERTATION.pdf

Top 5 countries by file download

	12 2015 - 06 2016
United States	26,964
Romania	91
Germany	82
South Korea	42
United Kingdom	38

Questions?