



# Batch Processes for Faculty Work

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

Texas Conference on Digital Libraries, May 25, 2016



# Texas ScholarWorks

- Created in 2008
- Consists mostly of library managed collections
- Content from all areas of the University
- Not a lot of faculty content



The screenshot shows the homepage of Texas ScholarWorks. At the top, there is a logo for The University of Texas at Austin, featuring a star and an open book. To the right of the logo, the text reads "The University of Texas at Austin" in orange, "Texas ScholarWorks" in dark blue, and "University of Texas Libraries" in a smaller, italicized font. Below the header, the text "Repository Home" is displayed. The main content area features three large images of the Austin skyline, each with a small white box containing a link label: "FAQ", "OA Blog", and "Policies". Below these images, the word "Introduction" is written in a large, bold font. At the bottom of the page, a paragraph of text states: "Texas ScholarWorks was established to provide open, online access to the products of the University's research and scholarship, to preserve these works for future generations, to promote new models of scholarly communication, and to help deepen community understanding of the value of higher education."



# Texas ScholarWorks

- Conference proceedings/journals: 159
- Student Work (not ETDs): 1,525
- UT Communities: 11,949
- ETDs: 17,983
- Faculty/Researcher Works: 2,609
- Total: 34,225



# Faculty publications



Hard to demonstrate our value without use



<https://hatponicsblog.wordpress.com/tag/children/>

<http://www.mackinacbridge.org/photo-gallery-10/>



# We Decided to Proactively Add Content

- Faculty/researcher published articles
- Where allowed by license (ex: 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?



<http://www.public-domain-image.com/free-images/people/children-kids/kids-look-for-birds-with-their-binoculars/attachment/kids-look-for-birds-with-their-binoculars>

# 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.

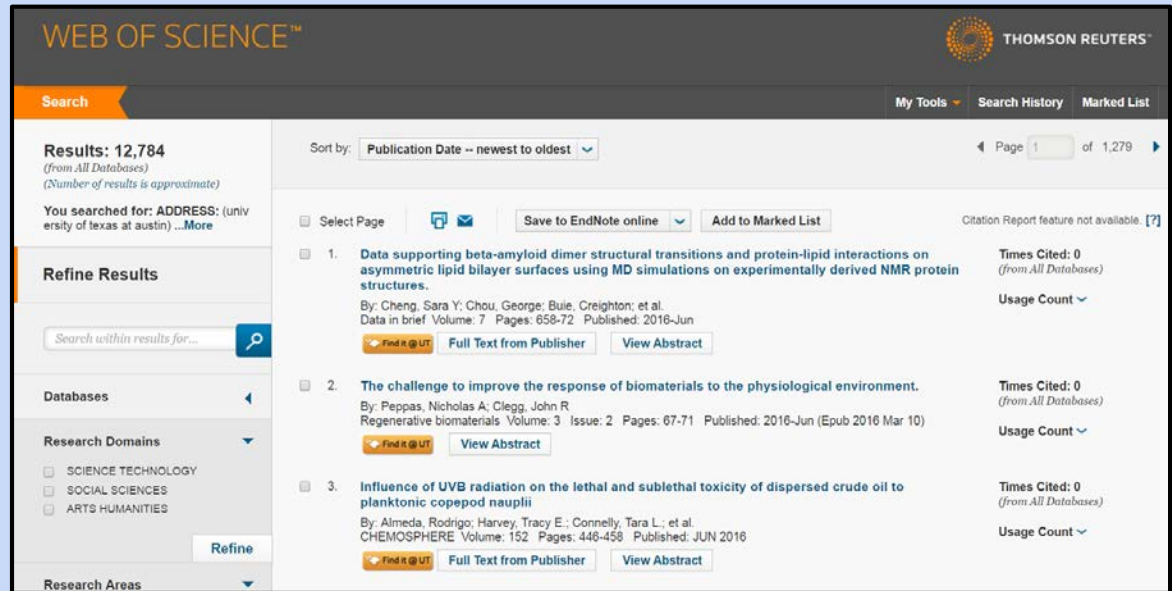


Journal:	<a href="#">Astrophysics</a> (ISSN: 0571-7256, EISSN: 1573-8191) [A translation of the Russian language journal]
RoMEO:	This is a <a href="#">RoMEO green</a> journal
Paid OA:	A paid open access option is <b>available</b> for this journal.
Author's Pre-print:	✓ author <b>can</b> archive pre-print (ie pre-refereeing)
Author's Post-print:	✓ author <b>can</b> archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF:	✗ author <b>cannot</b> archive publisher's version/PDF
General Conditions:	<ul style="list-style-type: none"><li>• Author's pre-print on pre-print servers such as arXiv.org</li><li>• Author's post-print on author's personal website immediately</li><li>• Author's post-print on any open access repository after 12 months after publication</li><li>• Publisher's version/PDF cannot be used</li><li>• Published source must be acknowledged</li><li>• Must link to publisher version</li><li>• Set phrase to accompany link to published version (see policy)</li><li>• Articles in some journals can be made Open Access on payment of additional charge</li></ul>
Mandated OA:	(Awaiting information)
Paid Open Access:	<a href="#">Open Choice</a>
Copyright:	<a href="#">Self-archiving policy</a> - <a href="#">Authors Rights</a> - <a href="#">Funder Compliance</a>
Updated:	16-May-2014 - <a href="#">Suggest an update for this record</a>
Link to this page:	<a href="http://www.sherpa.ac.uk/romeo/issn/0571-7256/">http://www.sherpa.ac.uk/romeo/issn/0571-7256/</a>
Published by:	<a href="#">Springer Verlag</a> (Germany) - <a href="#">Green Policies in RoMEO</a>



# 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



The screenshot displays the Web of Science search results interface. At the top, it shows 'WEB OF SCIENCE™' and 'THOMSON REUTERS™'. The search results are sorted by 'Publication Date -- newest to oldest'. The search criteria are 'ADDRESS: (university of texas at austin)'. The results list three articles:

- 1. Data supporting beta-amyloid dimer structural transitions and protein-lipid interactions on asymmetric lipid bilayer surfaces using MD simulations on experimentally derived NMR protein structures.**  
By: Cheng, Sara Y.; Chou, George; Buie, Creighton; et al.  
Data in brief Volume: 7 Pages: 658-72 Published: 2016-Jun  
Buttons: Find it @ UT, Full Text from Publisher, View Abstract
- 2. The challenge to improve the response of biomaterials to the physiological environment.**  
By: Pappas, Nicholas A.; Clegg, John R.  
Regenerative biomaterials Volume: 3 Issue: 2 Pages: 67-71 Published: 2016-Jun (Epub 2016 Mar 10)  
Buttons: Find it @ UT, View Abstract
- 3. Influence of UVB radiation on the lethal and sublethal toxicity of dispersed crude oil to planktonic copepod nauplii**  
By: Almeida, Rodrigo; Harvey, Tracy E.; Connelly, Tara L.; et al.  
CHEMOSPHERE Volume: 152 Pages: 446-458 Published: JUN 2016  
Buttons: Find it @ UT, Full Text from Publisher, View Abstract



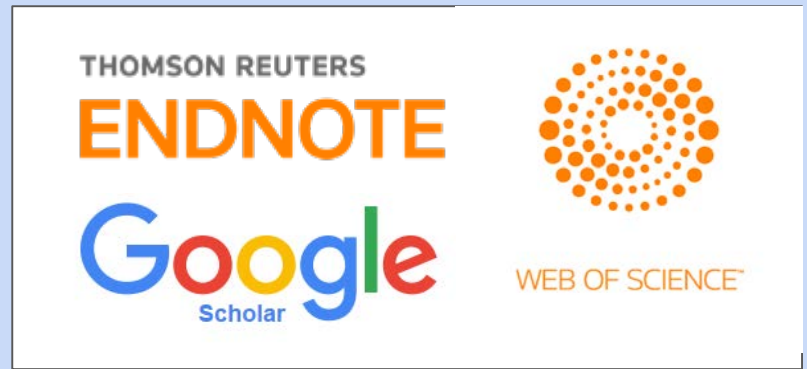
# Using Google Script to Automate Sherpa/Romeo Copyright Check

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

	F	I	J	K
	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	Embargo
	10736700	No publisher PDF	Final draft restricted	Embargo
	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	Embargo
ent	10860266	No publisher PDF	Final draft allowed	Embargo
oclimatolog	310182	No publisher PDF	Final draft allowed	no
	310239	No publisher PDF	Final draft restricted	Embargo
	8831351	Unknown	Final draft restricted	no
	314528	Unknown	Final draft allowed	no
ychology B	1461672	No publisher PDF	Final draft allowed	Embargo
	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:reated	dc.righ	dc.lang	dc.iden	dc.description	dc.identifier.citatic	dc.description.sponsc	dc.contribution.its.resti	dc.subject	dc.description.abstract			
185	2005_Hoffmeyer.p	In Vitro Analysis	Hoffmeyer, Micha	Article	Cancer Cell Inte	1475-286	2005	Adminis	English	10.1186/1475-286-14	Cellular and Molec	Hoffmeyer, Michaela F	NIH/NCI CA83957-01A1Ur	Hoffmeyer, Michaela	Open	oncology	Background: Inflammatory br	
186	2015_05_Aiken.pdf	Triggered Seisr	Aiken, Chastity  Zi	Article	Bulletin of the S	0037-110	2015-05	Adminis	English	10.1785/JCI150037.110	Geological Scienc	Aiken, Chastity, Jessic	National Science Foundati	Walter, Jacob L.	Open	non-volcanic tremor  3.0 tohoku	We conduct a systematic sea	
187	2015_09_Allshous	Lagrangian Ba	Allshouse, Michae	Article	Chaos	1054-150	2015-09	Adminis	English	10.1063/1.4998111	Nonlinear Dynamir	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	10.1063/1.4998111	Nonlinear Dynamir	Allshouse, Michael R.,	ONR N000141210665	Allshouse, Michael F	Open	lagrangian coherent structures	While more rigorous and soph	
183	2007_12_Askan.pc	Full Waveform	Askan, Aysegul  A	Article	Bulletin of the S	0037-110	2007-12	Adminis	English	10.1785/JCI150037.110	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	1934_09_Barkley.p	A Dynamical S	Barkley, Dwight  K	Article	Chaos	1054-150	1934-09	Adminis	English	10.1063/1.1016371	Nonlinear Dynamir	Barkley, Dwight, and lo	NSF DMS 92-06224, ECS	Barkley, Dwight	Open	mathematics, applied  physics, r	A simple system of five nonlin	
191	2012_12_Blazevski	Using Scatterin	Blazevski, Daniell	Article	Chaos	1054-150	2012-12	Adminis	English	10.1063/1.1330111	Mathematics	Blazevski, Daniel, and	NSF DMS0901389  Texas C	Blazevski, Daniell  Fra	Open	phase-space  mathematics, app	Scattering theory is a conven	
192	2012_03_Brager.pc	Impaired Dend	Brager, Darrin H.	Article	Cell Reports	2211-1247	2012-03	Adminis	English	10.1016/j.celrep.2012.03.011	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.celrep.2013.08.011	Cellular and Molec	Bowen, Cai, Jeong-Ho	CA132813, NCI NIH HHS  C	Lee, Jeong-Ho  Paul	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/JCI150037.110	Geophysics, Instit	Clark, Kate J., Bruce W	GNS  EQC Biennial Project	Wallace, Laura M.	Open	north-istand  hawkes bay  great c	At the southern Hikurangi ma	
195	2015_09_Corwin.p	Modeling Cour	Corwin, Lisa A.  Gr	Article	Cbe-Life Scienc	1931-7910	2015-03	Adminis	English	10.1187/cbe.1503001	Science and Math	Corwin, Lisa A., Mark u	Howard Hughes Medical In	Corwin, Lisa A.  Dolan	Open	genomics education partnership	Course-based undergraduate	
196	2015_12_Corwin.pc	The Laborator	Corwin, Lisa A.  Rv	Article	Cbe-Life Scienc	1931-7910	2015-12	Adminis	English	10.1187/cbe.1512001	Educational Psycol	Corwin, Lisa A., Christ	National Science Foundati	Corwin, Lisa A.  Runy	Open	undergraduate research experie	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

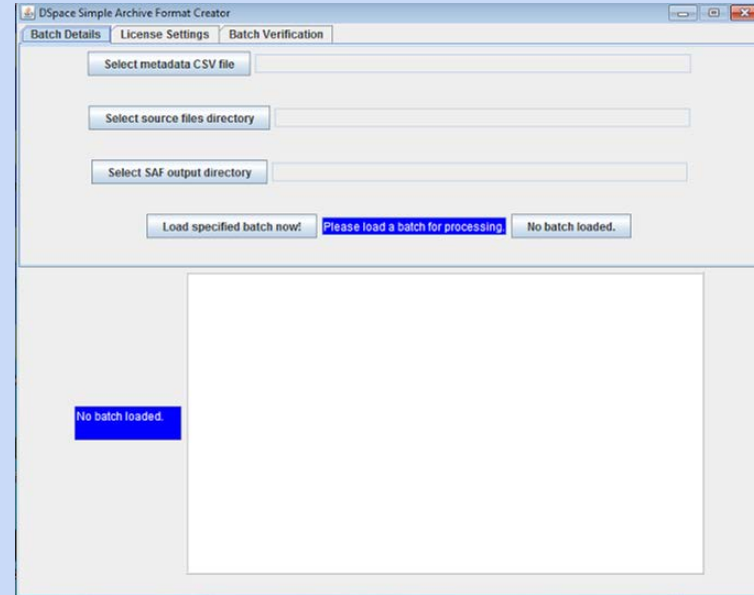


# Getting content into DSpace

- 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 2,609 items in Faculty/Research collection
  - Of these, 1,695 have been added in the last 18 months
- 718 waiting to be ingested
- 262,647 downloads since April 2014
- 0 negative comments





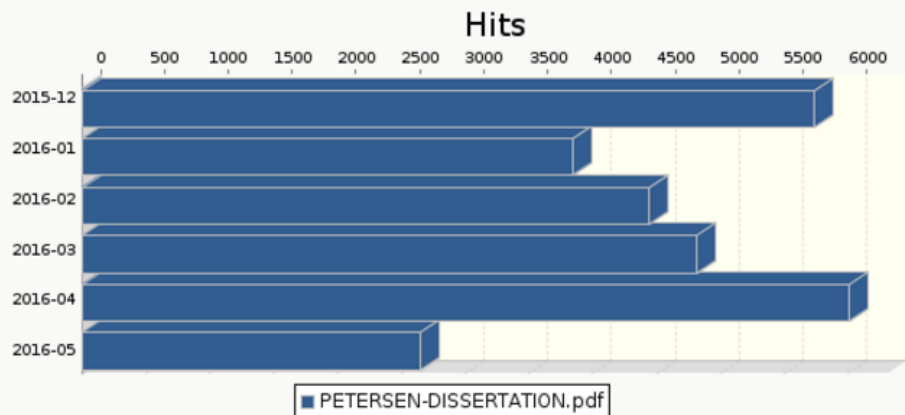
# Next Steps

- 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



## 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?

