

Purdue University
Purdue e-Pubs

Charleston Library Conference

Open Access, Open Access, How Does Your Catalog Grow? With Selection, Access, and Usage All in a Virtual Row!

David W. Schuster
Binghamton University, dschuste@binghamton.edu

Susan J. Martin
University of Chicago, smartin28@uchicago.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Collection Development and Management Commons](#), [Operational Research Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

David W. Schuster and Susan J. Martin, "Open Access, Open Access, How Does Your Catalog Grow? With Selection, Access, and Usage All in a Virtual Row!" (2016). *Proceedings of the Charleston Library Conference*.

<http://dx.doi.org/10.5703/1288284316458>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Open Access, Open Access, How Does Your Catalog Grow? With Selection, Access, and Usage All in a Virtual Row!

David W. Schuster, Director of Library Technology & Special Collections, Binghamton University

Susan J. Martin, Head of Acquisitions, University of Chicago

Abstract

Much of the open access (OA) focus and discussion has been on journals (think *Glossa*), but the open access monograph has come fully into its own. University and scholarly publishers are providing high-quality books, often in areas that rely on long-form scholarship. However, open access monographs presented a challenge. How do they fit into the traditional models of selection, acquisition, cataloging, and tracking usage?

In the spring of 2016, Texas Woman's University Libraries created a simple workflow to make open access monographs accessible through the libraries' discovery layer using Google Sheets to track the workflow and EZproxy to track usage.

Introduction

Texas Woman's University (TWU) is a public doctoral/research university and has campuses in Denton, Dallas, and Houston. With a student enrollment over 15,000, TWU is the nation's largest university primarily for women. It offers both traditional and online degrees in the liberal arts, education, business, nursing, health sciences, and the hard sciences. The TWU Libraries hold over 600,000 volumes, subscribe to over 2,000 journals and databases, and have a collections budget of \$1.7 million.

Texas Woman's University had begun initiatives to promote and include open access scholarship on campus and make data-driven decisions. As a leader and an integral part of that effort, the libraries were interested in supporting and including open access materials in the collections when possible. Throughout early 2016, open access initiatives filtered through our e-mail and into our conversations. However, lacking a process, these open access materials never moved beyond discussion and into our collection. We realized that we needed a systematic way to review and add these materials.

Formal meetings were not the norm in our library, but we decided that creating an open access workflow warranted one. Representatives from acquisitions/collection development, cataloging, electronic resources, and information technology

gathered together to design the process. The goal was simple: An open access workflow that was simple, effective, and flexible.

We decided to use Google Sheet as the basic workflow mechanism. Google would allow for multiple editors. It would track changes, allow comments, allow for multiple staff members to concurrently access the sheet, as well as provide notifications. The workflow would move from the selection process in collection development to cataloging for MARC record decisions, then to electronic resources for URL and proxy enabling. Information technology would become involved once a month to extract usage data.

How Do We Identify Open Access Materials?

The workflow was easy. To a certain extent, it mirrored the purchased workflows for electronic monographs. The challenge for collection development was identifying open access materials to acquire. In print collection development, selecting titles is easy. There are established tools and methods that notify selectors and libraries about new and forthcoming titles. However, not all publishers or book vendors incorporate open access materials into their catalogs, flyers, and selection tools. At TWU, we relied primarily on open access announcements on library listservs, at conferences, and from fellow librarians or university faculty.

How Do We Select?

Texas Woman's University uses the Ex Libris Alma system. One of the options Alma provides is a community zone of electronic journal and e-book packages. Libraries can easily "turn on" access to their subscription and purchased electronic book and journals packages. Additionally, Alma provides such community zone bibliographic records for the Directory of Open Access Journals (DOAJ) and the Directory of Open Access Books (DOAB), and the HathiTrust. When we first went live with Alma in 2012, we activated the DOAJ, DOAB, and HathiTrust collections, making available OA material through Primo, our discovery layer.

We revisited this decision during our OA discussions. Do we need to provide access to all these free materials? Perhaps it would be better to select based on our crafted collection development policy and allow researchers to discover any additional open materials via Google Scholar? We reversed our initial 2012 decision to make everything available and instead looked toward our collection development policy and criteria. If it did not fit into our collection plan, we were not going to add it even if the book was free, but having said that, we hoped that OA monographs might be used to supplement areas of the collection where budget constraints would not allow for acquiring materials of tertiary importance.

Our first open access monograph selection was the open access art book collection made available from the Metropolitan Museum of Art. This collection was selected based on the results of a LibQUAL survey the libraries conducted during the spring of 2015. Feedback indicated that faculty and students needed additional visual art and photography materials. Since we were not able to expand the visual art budget to substantially increase the collection with new print books, we looked toward open access to help alleviate some of the perceived deficiencies.

The Met materials were perfect. They were high quality and covered a variety of art disciplines and areas. By adding them, we would be able to supplement not only the visual arts department's needs but that of fashion design. Moreover, the books came with an added bonus: Free machine-readable cataloging (MARC) records. After adding the collection to Primo, we extracted selected ISBNs and highlighted these open access books on our new books widget, placed prominently on the libraries' homepage, to market our new additions and hopefully boost usage.

In addition to expanding the libraries' collection with supplemental materials, we decided to add access to an open access monograph if it could provide access to materials the libraries already owned in print but were not generally available to the public. For this, we targeted our children's historical collection (CHC).

Located on the Denton campus in the Blagg-Huey Library, the children's historical collection is a noncirculating collection of approximately 5,200 items dating primarily from the 19th and early 20th centuries. It includes picture books, fiction, classic children's series, poetry, nonfiction, as well as early readers, primers, and some textbooks. Many of the books are fragile and irreplaceable, and they are housed in the libraries' special collections vault. As part of an inventory and assessment of the collection, collection development searched the HathiTrust to see if any of the materials were openly available. If one of the CHC books was openly available, we added it to the open access spreadsheet. Since the HathiTrust is part of Alma's community zone, opening up these resources was easy: Locate the title, connect it to the institution zone, and update the proxy. We hope that by providing open access to materials from this collection, TWU faculty and students in education and library science would be able to more fully utilize this specialized resource in their research and coursework.

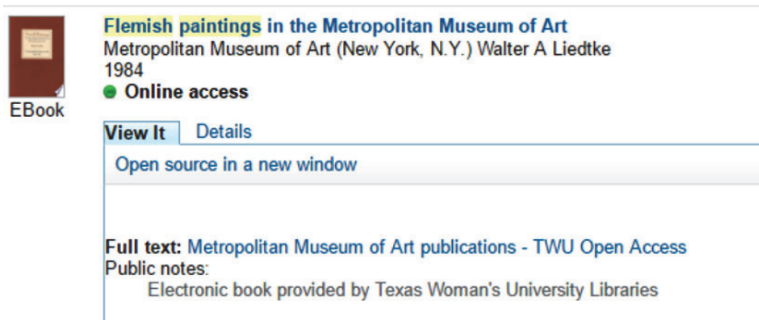


Figure 1. Example of Met title online.

In addition to content, a selection decision point was the e-book format. Publishers often make their open access monographs available in multiple e-formats, such as PDF, HTML, even Mobi and EPUB. We decided to make only one format available through our discovery layer, opting for the PDF version when available, since most browsers and operating systems have some type of Adobe viewer. We also opted to link to the actual item and not just the open access site. When a patron clicked through to the resource, the book would open in the viewer.

How Do We Manage Workflow?

Since open access materials did not need purchase orders for invoices, we opted to develop a simple process that was outside the purchased monograph (print and electronic) workflow. Instead of using Alma to manage the workflow, we decided to use a Google Sheet for interdepartmental workflow management. Google Sheet was flexible and allowed for customized fields, multiple users could work in it at the same time, and it tracked all modifications. By using the notification feature, each person along the work line knew when modifications and/or additions were made.

Once we began using this process, we were able to easily count the number of open access titles we were adding through our discovery layer. We started to track the number of open access e-books added into our monthly acquisitions statistics as a separate line item, which should help with Association of College

and Research Libraries (ACRL) statistical reporting. Since this workflow was new and experimental, the librarians from each area took ownership and did the work. Beginning in collection development, the acquisitions librarian entered in the data for each selection. The sheet contained the date requested, name of the open access collection, title, author, ISBN, the URL, and any additional information such as the availability of MARC records or Online Computer Library Center (OCLC) numbers.

Using the libraries' established standards for MARC records, the catalog librarian evaluated any freely available publisher provided MARC records and checked Alma's community zone to see if the resource was included. If the existing records did not meet TWU's cataloging standards, they would import records from OCLC. The electronic resources librarian verified the URL and added in the proxy in conjunction with information technology, who managed the library's EZproxy and maintained the appropriate stanzas.

One final workflow step was to market the resource. Each month, the acquisitions department would prepare a listing of newly received books to highlight using a LibraryThing widget. When open access title-by-title selection was implemented, acquisitions began adding in the open access monographs to its monthly lists. Now the new books widget contains purchased print, electronic, and open access items.

	A	B	C	D	E	F	G	H	I	J
1	Date Requested	Collection Name	Public Series Name	Title	Author	ISBN	Additional Information	Notes from Collections	Date Added into Alma	Notes from Cataloging
2	5/2/2016	Oxford University Press Open Access	Oxford Scholarship Online - TWU Open Access	Field trials of health interventions	Smith, Peter G.	978-198732880	https://global.oup.com/academic/doi/10.1093/acprof:oso/9781987328807/oa001		5/9/2016	
3	5/16/2016	Knowledge Unlatched Pilot Collection	Knowledge Unlatched - TWU Open Access	On Global Citizenship	Tully, James	9781849668162	http://www.knowledgeunlatched.org/ocollections/magnum-opal		5/17/2016	
4	6/1/2016	Knowledge Unlatched Round 2	Knowledge Unlatched - TWU Open Access	Midec Evangelicals	O'Connor, Mary I.	9781607324249	http://www.knowledgeunlatched.org/ocollections/		Coming soon	
5	6/10/16	Project Muse OA	Project Muse - TWU Open Access	The Self and its Passions: Betwixt, Between, and the History of the Disoriented Subject	Dean, Carolyn	9781501708403	http://muse.jhu.edu/about/oa_open_access_books.html and http://muse.jhu.edu/book/45091		6/17/2016	
6	6/15/16	luminos	Luminos - TWU Open Access	Migrating into Financial Markets	Bakker, Matt	978-0-520-94093-0	http://lib.oxj.org/0520940930		6/17/2016	
7	6/20/16	OpenBookPublishers	OpenBook Publishers - TWU Open Access	Oral literature in Africa	Finnegan, Ruth	978-1-906924-72-0	http://www.openbookpublishers.com/handle/17536/172	This book is accompanied by original recordings of stories and songs collected by the author; they are openly available here: http://www.oralheritage.org/actions/finnegan_001.html	6/21/16	
8	6/20/16	Jurix e-books	Purdue ePub - TWU Open Access	A Passion for Ideas: how innovators create the new and shape our world	Fleier, Heinrich	9781612491370	http://www.jurix.org/View/View.aspx?ID=1009&ContentID=5426		6/21/16	
9	6/4/16	University of Michigan		Campaign finance and political polarization: when Purists prevail	Le Raja, Raymond	978-0-472-90003-9	https://odod.lib.umich.edu/umdl/3344444.0001.001			
10	8/9/2016	Waltham		Recent advances in nutrition with particular reference to protein metabolism	Cannon, Paul R. B. 1892 (Paul Rother)		https://babel.hathitrust.org/cgi/pt?oclc=3344444.0001.001	1950 ed.		
11	8/9/2016	Waltham		The control of hunger in health and disease /	Cannon, Anton J. 1875-1956. (AMNH Author)		https://babel.hathitrust.org/cgi/pt?oclc=3344444.0001.001	1914		
12										
13										

Figure 2. Example of original Google Sheet with columns used.

How Do We Analyze Usage?

At TWU, the electronic resource librarian gathers and maintains, on a monthly basis, all database and e-book statistics for the libraries. These statistics are heavily used by collection development and acquisitions who rely on them for renewal decisions and collection analysis. While open access books would not need to be renewed, collection development was still interested in seeing usage. Usage would still help inform general collection decisions. Additionally, usage would also assist in evaluating the entire open access workflow. Is open access title-by-title selection worth it? Or should collection development encourage subject libraries to add open access monograph collections on their LibGuide pages, and rely on Google Scholar?

What we came up with was not perfect but a workable solution of using our proxy server to track usage at a basic level. However, since EZproxy log gathering does not fall under the electronic resources librarians' purview, information technology agreed to assist with this part. Sending a free resource through a proxy seems odd and counter to the open movement. However, we were interested in capturing usage and using the proxy was the immediate solution.

Since we decided to use EZproxy, we needed to use a log analyzer. At TWU we used Sawmill, but there are many different ones available, such as AWStats, Splunk, ezPAARSE, The Webalizer, and FastStats Log analyzer to name a few. As we started to explore the proxy log in Sawmill, we realized not enough of the URL was being captured, and our URLs in the spreadsheet were not always consistent. The workflow spreadsheet contained the URL to the monograph, but it did not track the final, proxied URL that underpinned the record in our instrument landing system (ILS). Without this specific URL, finding the various OA books would be near impossible, as our system logged well over 700,000 lines a day. Additionally, we discovered not enough information was being logged in the proxy to identify individual books at the same host site. We did not have a chance to test any additional system before we left for new opportunities.

Questions Raised and Improvements for the Future

Our fledgling process works, but it is far from perfect. The next step in this workflow evolution would be to fold open access monograph selection into the existing acquisitions and cataloging channels and improve and streamline usage statistics collecting.

TWU's ILS, Alma, is purchase-order-line driven, and the purchase order line is the beginning for all new electronic and print materials. Would it be worthwhile to create purchase orders for OA books and route them through the same channels as a purchased e-book would go? It would routinize the process, allowing ordering assistants in acquisitions to create and route the purchase order, and it would eliminate most of the need to maintain a separate spreadsheet. The exception is the usage statistics component. How would we keep track of the URLs to extract from the proxy logs?

Another issue that needs to be addressed is the long-term management of these open access resources: How will changes in URLs be tracked and managed?

We also need to be careful and consistent about our URLs. Consistent URLs for items within the same collection should be maintained rather than having the links for one e-book in a collection go to one provider and another link in the collection go to a different provider. We discovered that using MARC records from different sources often caused the root of the URL to be different.

Gathering usage statistics proved to be more difficult and problematic than we originally thought. Should we invest in another type of log analyzer for better usage tracking and evaluation? It may be worth investigating to see if another analyzer would allow us to easily gather information from a particular resource provider, thus improving our ability to calculate package statistics as well as individual e-book usage. Another desideratum is the ability of the log analyzer to export statistical data in COUNTER statistic formats. This would align the OA statistics with those from the libraries' fee based

databases and e-books and, thus, allow more productive comparisons. Could we automate the statistical harvesting process? We would like to have specific reports be designed to automatically run month to month. Would we get the full data? We were capturing all data possible within the proxy logs, but what we were extracting in the analyzer was only partial; it truncated the results in the analysis.

We would also like to revisit the default e-book version. What e-book versions do our patrons really find the most beneficial? We could analyze purchased e-book statistics to see if there are any patterns or preferences on e-book views or downloads. In addition, we could possibly work with our assessment staff to create a tool and survey our community. We need to be mindful of developments and changes in the university's course management system, as well as our researchers needs in terms of data mining and analysis.

We also discovered that some of our e-book aggregators were including open access monographs as part of their subscription collections. It inflates their collection numbers, but our statistics may not have been specifically from our catalog's links.

Conclusion

Data-driven decisions drive libraries, and it is crucial to be able to track and assess how libraries and their patrons interact with resources. Identifying those resources and bringing them into the catalog can be time consuming. The fruits of the labor is quality resources for the end user they may have never found. This article outlined our attempt to use open access freely available e-book monographs and track usage. We have since left Texas Woman's University Libraries for other gardens, and we have not been able to do any reassessment, exploration, or refinement to this process.