Using Institutional Repositories to Make Purchasing Decisions

Richard Wisneski

Marsha Miles
Cleveland State University, m.a.miles24@csuohio.edu

Follow this and additional works at: https://engagedscholarship.csuohio.edu/msl_facpub

How does access to this work benefit you? Let us know!

Original Citation

Repository Citation

This Contribution to Books is brought to you for free and open access by the Michael Schwartz Library at EngagedScholarship@CSU. It has been accepted for inclusion in Michael Schwartz Library Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact library.es@csuohio.edu.
Institutional repositories serve not only as a means to preserve and promote institutional faculty, student, and staff scholarly output but also as a means to assist a library’s acquisitions department in making informed, nuanced purchasing decisions and assessing current collections. After a brief review of some of the purposes of institutional repositories, we will explore access points within digital institutional repositories that assist with making informed acquisition decisions, particularly with regard to faculty and student publications and research interests and the analysis of repository statistical data.

LITERATURE REVIEW

While much has been published about institutional repositories themselves, less has been published about how acquisitions work can be informed by their contents. Hanson, Lightcap, and Miguez have written about the need for acquisitions departments to adapt to institutional repositories by understanding their metadata structures and utilizing acquisitions connections by standardizing acquisition-related metadata in institutional repositories.¹ Other authors have explored altmetrics, Open Access, and copyright issues in regard to institutional repositories. Bonilla-Calero, for example, looks at the ways one can examine scholarly output by a university through examining the contents of one’s institutional repository in addition to Web of Science and Scopus services.²
In 2008, Rick Anderson wrote that library trends showed growth in patron preference for e-resources, unique collections, and locally produced scholarship like that found in institutional repositories. As institutional repositories became more popular, challenges of implementing and maintaining them surfaced, including populating them and navigating permission and quality concerns. Morrow and Mower recommend increasing faculty awareness of scholarly communications issues and author rights. Other solutions to populating repositories with quality, permitted scholarship include marketing to specific disciplines, mediating deposits, and becoming data curators on campus. Wesolek suggests surveying end users to gain insight into who uses the materials and what type of content might be useful to them, which can influence marketing efforts and collection development.

Giescke stresses that repositories would not succeed if completely separated from other library functions instead of being part of digital content management departments and core library services. As libraries evolved and experienced a shift from print to electronic journals and fewer book purchases, Douglas and Flinchbaugh pointed to transferrable serials and acquisitions staff skills that could be leveraged to complete institutional repository–related work. Rossmann and Arlitsch write about the need for libraries to shift from budgeting for the purchase of materials to the delivery of materials based on priorities of their users. Delivery and access mechanisms include integrated library systems, discovery layers, and institutional repositories.

According to David Lewis, academic libraries would also need to change their collecting practices to focus on e-resources, on-demand purchasing and subscriptions, unique local materials (like those included in institutional repositories), and Open Access journals. Kumar and Dora analyzed citations from dissertations completed at the Indian Institute of Management to make informed collection management decisions. A study by Hoskins found that Open Access initiatives at South African universities did not significantly influence journal cancellations.

As education transforms to a more open model, there are an increasing number of open educational resources initiatives worldwide. When Yang and Li surveyed faculty, they found that most respondents were aware of Open Access journals in their fields and willing to publish in Open Access publications; however, many were
unaware of the process to contribute to the institutional repository.\textsuperscript{15} Gaines promoted discipline-specific Open Access discussions with faculty to address their concerns and meet their specific needs.\textsuperscript{16}

Howard points out that new ways of measuring scholarly influence are being explored, specifically altmetrics, which measure scholarly interactions online.\textsuperscript{17} Galligan and Dyas-Correia explain that altmetrics and Open Access publishing have gone hand in hand. One example they mention is using altmetrics for publications in Open Access institutional repositories that may not have sophisticated metrics available.\textsuperscript{18} Konkiel and Scherer wrote on the benefits for authors, repositories, and university administrators.\textsuperscript{19} As important as these various studies are, there is nonetheless room for exploring how institutional repositories can also inform acquisitions librarians to obtain material and evaluate current library holdings.

**CLEVELAND STATE UNIVERSITY’S INSTITUTIONAL REPOSITORY**

Located in downtown Cleveland, Ohio, Cleveland State University (CSU) has an enrollment of more than 17,000 students and approximately 580 faculty members. In order to showcase and preserve faculty, staff, and student scholarship and creative works, the Michael Schwartz Library launched CSU’s institutional repository using bepress Digital Commons in March 2012.

Today, our repository includes more than 15,000 papers in over 700 disciplines. The repository hosts a variety of materials including books, conference proceedings, journals, images, videos, oral histories, and open educational resources. A number of the books and journals in the repository were published with the imprint of the Michael Schwartz Library, MSL Academic Endeavors.

The Michael Schwartz Library staff includes nine liaison librarians, and their repository work assists them in becoming more familiar with their faculties’ scholarship and research interests. They also help their faculty create professional profile pages using a companion program of Digital Commons, SelectedWorks.

Liaison librarians are responsible for collection development in their respective subject areas. They draw upon their knowledge of faculty research interests and student needs in making decisions to
acquire, cancel, and retain material. They also make use of quantitative data from COUNTER-compliant statistics, consortial data, and financial data generated from our subscription agent and integrated library system (ILS). For example, subject librarians make use of interlibrary loan and consortial borrowing data to see what has been requested by our patrons as one means to determine what to acquire. The institutional repository can provide another means to investigate what is being used by patrons for future acquisition decisions.

The Michael Schwartz Library fosters collaboration among its staff. With shrinking budgets and staffing, collaboration is even more important. We are continually pursuing ways to become more efficient, eliminate duplication of efforts, and do more with less while maintaining excellent service. Library systems staff are willing to assist with projects and make workflows more efficient to benefit the library and the campus community. They are integral in the implementation of some of the following workflows.

**ACCESS POINT: CITATIONS IN FACULTY PUBLICATIONS, THESIS, AND DISSERTATIONS**

An institutional repository provides several access points acquisitions librarians can take advantage of to better inform their work. Whether a homegrown or commercial product, institutional repositories typically include publications by faculty and staff and theses and dissertations by graduate students. These publications contain three key access points: journal titles, references/works cited pages, and keywords/subject headings.

Journal titles are a quick means for acquisitions librarians to collect information to see whether the library subscribes to the title. Depending on one’s institutional repository, one can grab these titles via a record’s metadata and export them into a spreadsheet to compare what journals authors are publishing in with whether the titles are available in one’s library.

Another data point comes by exploring the references from the publications themselves. There are different methods to extract this information, depending on one’s institutional repository and available data mining resources. In some instances, an SQL query in a
homegrown database system can extract citations in publications. Another method involves screen-scraping and exporting data in XML or raw text format. Using CERMINE, for example, one can parse digital object identifiers (DOIs) when available. One can also use Elasticsearch, with sister tool Kibana, to parse and visualize the data. For instance, one can parse <back> nodes and child nodes from publications into one file. Thus, one can capture the <source> and <article-title> information from publications:

<source>Journal of Advances in Modeling Earth Systems</source>
<article-title>Transportation network analysis</article-title>

Concatenating this information into one file allows an acquisitions librarian to then run this list against current library holdings to see what publications faculty and students are citing, to what the library subscribes, and coverage of holdings. In making purchasing decisions, subject librarians can use this data to see whether there are subscriptions or monograph holdings that the library should purchase.

Similarly, acquisitions librarians can also take advantage of theses and dissertations ingested into their institutional repository. Once again, depending on the repository an institution is using, scripts can be run to text-mine the back matter from theses and dissertations to identify what publications are being cited and then run this information against holdings in an ILS or knowledge base. Running XSLT and generating compound XML files for back matter from references and works cited pages, acquisitions librarians can again see what works authors have cited that the library does not subscribe to or own.

**ACCESS POINT: SUBJECT HEADINGS AND KEYWORDS**

Author-provided subject headings and keywords provide another data access point. In our institutional repository, author-supplied keywords and subject headings are found in the metadata for each thesis and dissertation. Our bepress Digital Commons allows us, via its Dashboard, to export subject headings and keywords into an Excel file for further analysis. Alternatively, with assistance from a systems
library, we use screen-scraper technology that grabs dissertations’ or theses’ metadata subject elements collectively and outputs a CSV file for each dissertation or thesis (see figure 20.1).

We then compare these subject headings to those utilized in our ILS and knowledge base to identify matches and unique headings. The latter is especially helpful if there are subjects in which we have deficient holdings.

Author-supplied keywords or those supplied by the publication itself provide another access point. Depending on the institutional repository being used, both data points can be recorded in a separate file, such as an Excel file, at the point of ingestion, or later from downloading citation data from the whole repository via an institutional repository–supplied dashboard, database query, or screen-scraping programs, or a screen-scraping run against an article or publication’s listing in the institutional repository when none of the other options are available.

Keywords provide an access point to possibly identify what subjects institutional authors are identifying for their research purposes. Acquisitions librarians can use this information to run queries in the ILS to see how well such subjects are covered.

In the absence of controlled vocabulary, keywords can be analyzed in terms of frequency and related discipline. For example, Kibana allows for data visualization (e.g., word clouds) to see word frequency. This information can then be run against a library’s physical or electronic holdings to identify how strong the collections are in those areas.

### ACCESS POINT: BOOKS

Cleveland State University’s repository currently includes over 300 books, including over 100 books in the faculty scholarship collection. These books were written by CSU faculty focusing on their diverse research interests. A number of the e-books focusing on the history of
greater Cleveland and Northeast Ohio have been digitized from origi-
inals held by the Michael Schwartz Library Special Collections, and a
few were born digital. There are also collections for CSU alumni pub-
lications, books for sale by the library, and books published through
the Michael Schwartz Library Academic Endeavors.

However, the full-text is not available for all books in the reposi-
tory’s collections. In these cases, a link to the full-text or “find at the
library” is included. Books might be purchased based on high meta-
data page hits if they are not already owned by the Michael Schwartz
Library. This data can be downloaded as an Excel spreadsheet through
our institutional repository Digital Commons Dashboard.

**ACCESS POINT: FACULTY EXPERTISE**

As mentioned earlier, CSU subscribes to bepress’s SelectedWorks to
create professional profile pages for faculty, including their areas of
research, expertise, or research interests. Many institutional reposi-
tories in general typically include a faculty expertise component. Again,
various means can be implemented to capture and export this data,
depending on the infrastructure of one’s institutional repository. For
CSU, reports including this information can be generated using the
bepress Dashboard (see figure 20.2). For example, we can identify
that one faculty member’s research interest is “narrative realism.”
Keyword or subject searches in an ILS or knowledge base can show
how extensive and current a library’s holdings are in these areas.

![Figure 20.2](image.png)

**Figure 20.2** This image shows a portion of our bepress institutional repository dashboard. This
dashboard also provides a means to download specific titles to Excel to analyze downloads and
usage in more depth.
FURTHER INSTITUTIONAL REPOSITORY ANALYTICS

Analytics either provided by or pulled locally from an institutional repository as a whole can inform acquisitions librarians on such matters as what schools and departments produce the most scholarship, trends over time in scholarly output, and usage data.

Through a request sent to our institutional repository vendor, bepress Digital Commons, we obtain quarterly statistics customized according to our specifications. These customizations contain data in an Excel worksheet (see figure 20.3), including the following:

- Document type
- Original journal/publication, including citation information (date, volume, issue, pagination)
- Discipline(s)
- All-time downloads
- All-time page hits

Our institutional repository also provides a means to see at any time usage statistics via a data dashboard. One can filter by specific titles, departments, and schools, as well as adjust the date range (see figure 20.4).

Both statistical reports have their particular uses. In the former, we can do further data analysis in Excel to see what works, authors, and departments have the most usage, and trends over time. We can also filter specific journal publications and disciplines (see figure 20.5).

As with identifying what publications authors are citing in their works, we can use this data to see what publications and presses authors are publishing in and whether we have access to them. If not, particularly for those that get the most use, we can use this information in making purchasing decisions.

In the latter usage report obtained from bepress Digital Commons, we can see in graphical representations use over time for particular works, or within departments and colleges (see figure 20.6). These reports can be exported to Excel for further analysis, including trends over time and comparisons between departments.
**Figure 20.3** Abridged view of data supplied by bepress Digital Commons.

**Figure 20.4** Snapshot from bepress Digital Commons usage statistics.

<table>
<thead>
<tr>
<th>All-time Downloads</th>
<th>Disciplines</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Grand Total</td>
<td>(Multiple Items)</td>
</tr>
</tbody>
</table>

**Row Labels**

- Brigham Young University Law Review
- Brooklyn Law Review
- Capital University Law Review
- Cleveland State Law Review
- Georgia Law Review
- Loyola Law Review
- Nebraska Law Review
- North Carolina Law Review
- Sociological Quarterly
- Texas Journal on Civil Liberties & Civil Rights
- University of Toledo Law Review
- (blank)

**Figure 20.5** Snapshot of top journals used in law.
In doing this kind of analysis, we take into consideration the number of assets in each school or discipline in proportion to the number of downloads and hits to account for variations in the number of assets departments submit to the repository. This analysis gives a fairer account of how much a school’s work is being accessed and may be especially helpful if libraries are dealing with stressed budgets. In other words, if this analysis shows that one school or program has had more access than another, then Acquisitions may determine that more monies should be invested in those that are more heavily used.

Data analysis of usage on specific institutional members’ works can be insightful for acquisitions decisions (see figure 20.7). Here, we’re not concerned with individual names but rather with rank and department affiliation. If the data show that some departments or ranks are more represented than others, this may tell us that there needs to be further outreach toward those departments and ranks that are underrepresented. Outreach efforts are important for Acquisitions
to shed further light on whether there needs to be more collaboration with certain departments or programs to promote, store, and distribute their work and solicit feedback from them on their collections needs.

**OPEN EDUCATIONAL RESOURCES IN INSTITUTIONAL REPOSITORIES**

Open educational resources are “teaching, learning or research materials that are in the public domain or released with an intellectual property license that allows for free use, adaptation, and distribution.”

With the rising costs of tuition and textbooks, open educational resources are becoming increasingly important.

In February of 2014, the then president of the Cleveland State University Student Government Association (SGA) talked about textbooks at a Faculty Senate meeting. Students wanted greater consistency
across a course and all textbooks on course reserve in the library. The provost also discussed textbook costs and asked for options. Then, the SGA executive board called for standardized textbooks across sections (such as math and other general education courses).

That April, the Student Life Committee annual report included exorbitant prices for required textbooks, how prices were rising greater than inflation, the idea for incentives for professors to submit orders to the bookstore in a timely fashion, and renting textbooks as an alternative to purchasing them. These topics were discussed but no charges or specific instructions were given at the time.

Even though CSU doesn’t have a formal Open Access policy statement, Open Access helps the library support students. Our director made the investment of time and energy in Open Access as a strategic initiative for the library to better serve students and faculty. This translated into goals for librarians and new collaborations within the university.

We have a small but growing collection of open educational resources created by CSU faculty in our institutional repository, including a virtual workbook for our Introduction to Geography course and several physics lectures. We also link to a few reputable open educational resources collections to help faculty get started when searching for quality open textbooks.

We link to OpenStax College (https://openstax.org/about), an initiative of Rice University, that hosts a collection of high-quality, peer-reviewed textbooks. They advertise them as professional quality textbooks that meet standard scope and sequence requirements. Faculty can customize them as needed for their specific course.

CSU is a member of the Open Textbook Network (https://research.cehd.umn.edu/otn/) which sponsors the Open Textbook Library (https://open.umn.edu/opentextbooks/). These textbooks have been reviewed by faculty from a variety of colleges and universities to assess their quality. They can be downloaded for no cost or printed at low cost.

We also link to bepress’s Teaching Commons (http://teachingcommons.us/). Teaching Commons brings together high-quality open educational resources that are curated by librarians and their institutions and includes Open Access textbooks, course materials, lesson plans, multimedia, and more. This user-friendly collection can be browsed by type of resource or subject area.
Another way we are promoting open educational resources is through textbook affordability small grants. With the provost’s support, the library has partnered with the Center for eLearning, the Center for Instructional Technology and Distance Learning, and the Center for Faculty Excellence to offer grants to CSU faculty to adopt low- or no-cost course materials. Faculty can adopt an existing open textbook; review an existing open textbook found in the Open Textbook Library; revise, remix, or adapt an existing textbook or open educational resource to enable a fully open course; or make use of materials that are Open Access or licensed through the library. They receive half of the funding up front and the balance when they complete the project and provide a report evaluating the impact on student learning at the end of the course.

Initiatives such as these are not unique to CSU. Use and acceptance of open educational resources is increasing, and many libraries are already involved with open educational resources initiatives on campus.21

OPEN EDUCATIONAL RESOURCES: CONNECTIVITY TO ACQUISITIONS

When conducting research or navigating copyright while revising, remixing, or writing their own open educational resources, faculty have the support of their liaison librarians. From this, acquisitions librarians can learn what research areas and software packages are needed for editing existing or creating new open educational resources.

Traditionally, libraries have not purchased textbooks to add to their collections. The Michael Schwartz Library has recently partnered with the bookstore to provide access to e-books owned by the library that are assigned for current courses and has also created a new Textbook Center offering a limited selection of print textbooks required for some general education classes.

These initiatives help inform acquisitions librarians as to which e-book packages to subscribe to or purchase. Acquisition librarians can also take open educational resources into consideration by keeping up-to-date on current and forthcoming open educational resources in the institutional repository.
CONCLUSION

A university’s institutional repository provides many access points to mine data and, in so doing, assists in acquiring new content. For example, an institutional repository provides information on where faculty and students are publishing their scholarly work, what sources they are citing, and what their areas of interest are by way of the expertise of keywords and subject identifiers from their scholarly work. These access points further inform acquisition analysis and decisions. Libraries can use open source tools and homegrown scripts, in addition to statistical reports provided by some commercial products, to obtain such information and run this data against current holdings. Of course, it is recommended that thorough project planning be undertaken as early as possible in the process to identify what specific access points a university wants to have in its institutional repository and what processes would be involved in gathering data from these access points. The information found in institutional repositories can be invaluable in providing materials and resources that faculty and students use and rely on for their scholarly work.

ACKNOWLEDGMENTS

Special thanks go to Justin Grogan-Myers for his technical assistance and support of CERMINE, Elasticsearch, and Kiban.

NOTES


