The Collections PBI: Interactive Data Visualization for Campus Outreach

Nat Gustafson-Sundell
Presenter

Evan Rusch
Presenter

Pat Lienemann
Presenter

Jeff Rosamond
Presenter

Abstract

The Minnesota State University, Mankato Collection Management Technology team presented a collection analysis and data visualization solution called the Collections Power BI (CPBI). Pat Lienemann provided an overview of the CPBI and a live demonstration. Nat Gustafson-Sundell discussed the development of the CPBI and aspects of the implementation. Evan Rusch described four example use cases for the CPBI and outlined future directions for development.

Keywords: collection outreach, accreditation, collection analysis, collection development, journal, data visualization

https://doi.org/10.3998/nasig.6731

Introduction

The Minnesota State University, Mankato (MNSU) Collection Management Technology (CMT) team presented on a journal collection analysis and data visualization solution the authors developed and implemented using Microsoft (MS) Power BI. This solution, called the Collections Power BI (CPBI), can be used across campus by librarians, faculty, and administrators to understand the value journals provide to academic programs across a range of subjects, collections or publishers, and numerous other variables. The CPBI combines journal data from about a dozen sources and presents the results as interactive charts and matrices.

Overview & Demonstration

Lienemann's overview of the CPBI covered the reports, data, and interface of the online service, followed by a live demonstration. In the current version of the CPBI, which is available across the MNSU campus to users as an online service, there are nine pages, or "reports." Each report includes one or more data visualizations. There are two types of reports: collection profiles (CP) and journal lists (JL). The collection profiles allow users to view the collection as a whole, or through any filter(s), based on aggregated data. The journal lists are matrices describing individual journals using numerous journal-level variables. Both types of report can be updated dynamically using filters. The journal lists can also be sorted based on any variable in the matrix.

The reports include data which can be divided into three main categories: quality (Q), supply (S), and usage (U). A fourth category of cost (C) data is used more rarely.

Quality variables, which include citation-based measures and ranks, are derived primarily from ScImago.[1] ScImago also provides other useful data, especially journal subject(s), as well as some supply

information, such as the number of citable documents published in any given journal over the past three years.

Journal supply variables are derived primarily from the Alma library management system (LMS). MNSU uses both electronic and print journal holdings and coverage data from Alma, then processes the data so holdings can be analyzed based on numerous categories. These categories include Any Access, Recent Access, and Current Access, as well as holdings types, such as Subscription, Non-Subscription, Open Access, Consortial Access, and Post-Cancellation Access. All of these categories can be summarized in the aggregate, such as the total number of subscription journals, or the total number of Open Access journals for any filter context, such as subject. The holdings types also include coverage information only meaningful at the journal-level, such as the coverage provided by the subscription or the coverage available as post-cancellation access.

Usage variables are derived from Counting Online Usage of NeTworked Electronic Resources (COUNTER) reports, as well as link resolver, print browse, and interlibrary loan (ILL) usage data, including any calculated measures based on these data. Calculated measures can include varieties of cost-per-usage calculations, but also home-grown measures. For example, MNSU has developed a "Southworth Ratio," named after the librarian who invented it, to calculate usage trends for journals or journal packages over time. In its current iteration, the Southworth Ratio is calculated by dividing the most recent four years of article downloads by the past eight years of article downloads. If the Southworth Ratio is below .5, then usage is trending down, and vice versa.

Filters provide the power of the CPBI online service interface. These can be used to update the reports dynamically. The available filters vary with each report, but they typically include subject, Library of Congress Call Number, source of access, ScImago Journal Rank (SJR) quality slicers, and a simplified version of the holdings type. These filters can be applied in any combination. For more information,

including further discussion of the data categories, Lienemann directed the audience to the MNSU Collection Analysis library guide.[2]

Lienemann provided a live demonstration of four reports. The first report's title "CP Subject QSU" indicates it is a collection profile report intended for use at the subject level, including quality, supply, and usage variables. The report includes a matrix and a column chart. The matrix shows journal data aggregated based on subject quartile. Variables include the number of journals available for the subject, the number of these journals for which MNSU provides any, recent, or current access, other journal supply information based on selected holdings types, the total of article downloads for the past nine years, the total of link resolver clicks for the past four years, the total of ILL requests for the past four years, the aggregated Southworth Ratio, and a cost-per-use ratio. The column chart graphically illustrates journal supply based on a selection of variables from the matrix. For each quartile, there are four columns describing the number of journals available for the subject, as well as the number of journals for which MNSU provides any, recent, or current access. Lienemann updated the subject filter to view soil science journals only (see Figure 1).

The second report's title "JL Subject QSU" indicates it is a journal list report intended for use at the subject level, including quality, supply, and usage variables. The report consists of one matrix only,

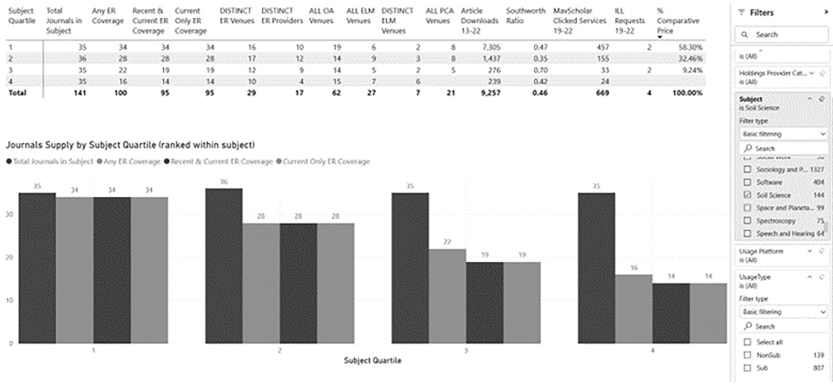
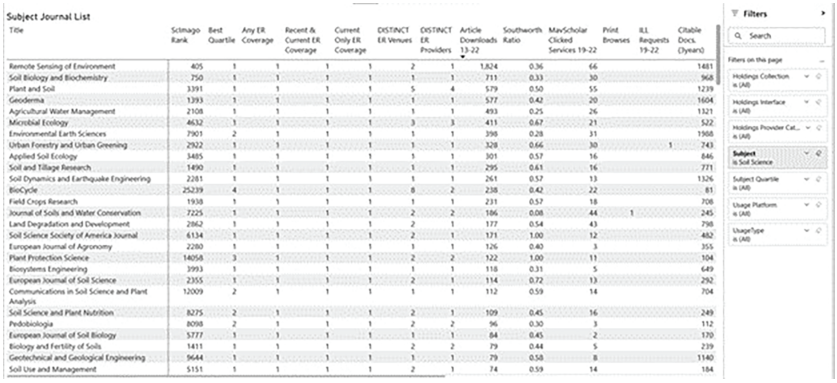


Figure 1. CP Subject QSU Report

listing all the journals and selected journal-level variables for a given subject. The variables include all the same variables as the CP Subject QSU report, as well as journal rank, best quartile, the total of any print browses for the past four years, the total of citable documents published in each journal in the past three years, the total of citations of articles in the journal over the same period, two citation-based metrics, the vendor, and the publisher of the journal. Lienemann updated the subject filter to view soil science journals only, then he sorted the list based on article downloads to show sorting functionality (see Figure 2).

The third report, "Journal Lookup," can be used to search individual journals and includes two line-charts, a combined line and column chart, and a matrix. The first line chart shows the article download trend over nine years for a given journal broken out by subscription and non-subscription usage. The second line chart shows the SJR trend over five years. The combined line and column chart shows the link resolver usage trend over four years, expressed as a line against columns showing the ratio of link resolver usage divided by article downloads. This ratio can provide a sense of how the discovery layer has been used to find the journal. The matrix provides journal information, including International Standard Serial Number (ISSN), best quartile, publisher, vendor, an overall summary of coverage from all



Title	Scimago Rank	Best Quartile	Any ER Coverage	Recent & Current ER Coverage	Current Only ER Coverage	DISTINCT ER Venues	DISTINCT ER Providers	Article Downloads 13-22	Southworth Ratio	MavScholar Clicked Services 19-22	Print Browses	ILL Requests 19-22	Citable Docs. (3years)
Remote Sensing of Environment	405	1	1	1	1	2	1	1,824	0.36	66			1481
Soil Biology and Biochemistry	750	1	1	1	1	1	1	711	0.33	30			968
Plant and Soil	3391	1	1	1	1	5	4	579	0.50	55			1239
Geoderma	1393	1	1	1	1	1	1	577	0.42	20			1604
Agricultural Water Management	2108	1	1	1	1	1	1	493	0.25	26			1321
Microbial Ecology	4632	1	1	1	1	3	3	411	0.67	21			522
Environmental Earth Sciences	7901	2	1	1	1	1	1	398	0.28	31			1988
Urban Forestry and Urban Greening	2922	1	1	1	1	1	1	328	0.66	30		1	743
Applied Soil Ecology	3485	1	1	1	1	1	1	301	0.57	16			846
Soil and Tillage Research	1490	1	1	1	1	1	1	295	0.63	16			771
Soil Dynamics and Earthquake Engineering	2281	1	1	1	1	1	1	261	0.57	13			1326
BioCycle	25239	4	1	1	1	8	2	238	0.42	22			81
Field Crops Research	1938	1	1	1	1	1	1	231	0.57	18			708
Journal of Soils and Water Conservation	7225	1	1	1	1	2	2	186	0.08	44	1		245
Land Degradation and Development	2862	1	1	1	1	2	1	177	0.54	43			798
Soil Science Society of America Journal	6134	1	1	1	1	2	1	171	1.00	12			482
European Journal of Agronomy	2280	1	1	1	1	1	1	126	0.40	3			355
Plant Protection Science	14058	3	1	1	1	2	2	122	1.00	11			104
Biosystems Engineering	3993	1	1	1	1	1	1	118	0.31	5			649
European Journal of Soil Science	2355	1	1	1	1	2	1	114	0.72	13			292
Communications in Soil Science and Plant Analysis	12009	2	1	1	1	1	1	112	0.59	14			704
Soil Science and Plant Nutrition	8275	2	1	1	1	2	1	109	0.45	16			249
Pedobiologia	8098	2	1	1	1	2	2	96	0.30	3			112
European Journal of Soil Biology	5777	1	1	1	1	1	1	84	0.45	2			170
Biology and Fertility of Soils	1411	1	1	1	1	2	2	79	0.44	5			239
Geotechnical and Geological Engineering	9644	1	1	1	1	1	1	79	0.58	8			1140
Soil Use and Management	5151	1	1	1	1	2	1	74	0.59	14			184

Figure 2. JL Subject QSU

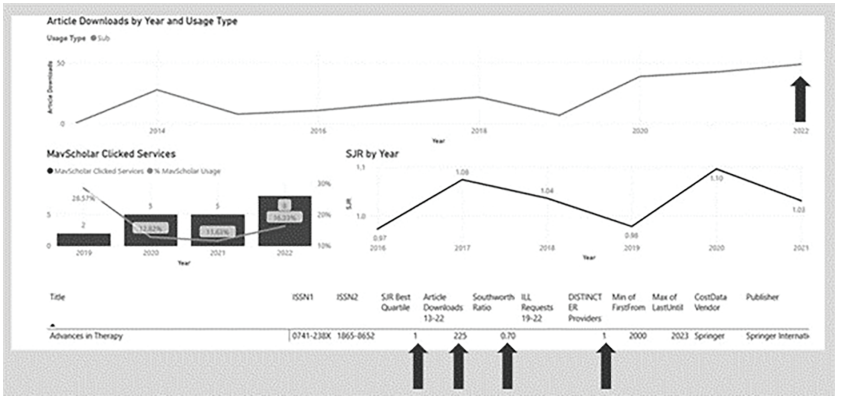


Figure 3. Journal Lookup

sources, a count of distinct online access providers, the total of article downloads for the past nine years, the total of ILL requests for the past four years, and the Southworth Ratio (see Figure 3 later in this paper).

Finally, Lienemann briefly demonstrated a report entitled "JL Access Changes," a journal list report intended for tracking changes to journal access. This report includes a matrix listing journals, the overall ScImago rank of each journal, and an access indicator (1 for yes or 0 for no) for each journal for each of the past four years (2019–2022). Lienemann updated the filter to view soil science journals only, demonstrating how the library can track changes to journal access in a subject area that is important to MNSU. A subject liaison might use this report for collection development purposes, for example, if MNSU has lost access to a journal previously provided only via an aggregator, or to communicate with a department about changes to access.

Background & Implementation

Gustafson-Sundell described the components of MS Power BI, why MNSU now prefers Power BI above other data visualization solutions, and how to implement Power BI, with a focus on data design.

Developers create reports in Power BI using a desktop application, while users access the reports through the online service. The desktop application consists of three views. These views provide different development environments and functionality. The data view can be used to manipulate the underlying data and to create new measures. For example, given usage (of any type) and cost, one could create cost-per-usage measures. Or given historical usage data over enough years, one could calculate a version of the Southworth Ratio. If these measures are created within Power BI, they will update with the filter context. The model view can be used to create and revise data relationships. Finally, the report view can be used to create data visualizations. Gustafson-Sundell commented that novice developers could get started using the report view primarily and just one or two tables of data. He thinks, however, the power in Power BI comes from the ability to combine many tables in the model view and to create measures in the data view.

To explain why MNSU currently prefers Power BI, Gustafson-Sundell provided a brief history of the CMT team's work on collection analysis. The team has iteratively developed collection analysis tools and methods over the past eight or so years. Early on, the team focused on how to increase the efficiency of data combination. Because the team was able to reduce the work required to combine data drastically, the team could combine more and more data for reports. To make these reports more legible and draw attention to interesting information, the team began developing data visualizations (viz) and to build more user-friendly, finished reports based on these data viz. The team first developed finished reports using Excel, then Tableau, then Python. In Python, they automated data viz production, including dozens of charts as well as report production.

Throughout these iterations, MNSU used the reports for a variety of applications successfully, including collection development and for accreditation, but they realized the reports would have more impact if more people could interact with them. They focused increasingly on how to develop interactive reports, first in Excel, then in Power

BI. Power BI is used across the MNSU campus to analyze institutional research data. University administrators, department chairs, and others use Power BI to extract information about enrollments, completion, and so on, so there is already a user base including the best possible target audience. The Power BI interface is interactive, device agnostic, and fairly intuitive. Power BI also improves production efficiency. By implementing Power BI, the CMT team was able to take advantage of a new approach to data design, reducing the number of steps to create reports. Power BI provides clear advantages over the team's previous solutions in Excel, Tableau, and Python.

Gustafson-Sundell mentioned there would not be time enough in the presentation to demonstrate the data processing steps, but he referred the audience to instructions.[3] The essential concern is to match any number of journal data sources to a "key list." A key list provides the central connection between any number of match lists. To combine data in Power BI, Microsoft encourages developers to use a "star schema" as a data model. In a star schema, there is one table at the center of the star serving as the junction for other tables, which are the points of the star. A star schema is just like a key list surrounded by match lists. In other words, any library could develop reports of the same depth and breadth as MNSU as easily by using the same or similar methods.

Gustafson-Sundell moved on to describe some problems encountered by the team in early versions of the implementation. Because the team could already combine numerous data sources, they did not start simply. They started with a large number of data sources and complex interrelationships in the data. It has taken time to get a handle on how the reports could yield false or confusing information in some circumstances. Filters can be especially problematic, depending on the filter and how the data sources are related. Gustafson-Sundell provided two examples, the first showing how results could be overfiltered accidentally, and the second showing how results could be multiplied accidentally. The team has simplified the data relationships and reduced the number of filters so that users can interact with the

data with more assurance of accurate results. In some cases, they have made the decision to continue providing filters that could yield false results if used incorrectly, but they have added instructions within the reports to reduce this risk.

Gustafson-Sundell concluded by saying it is easy to get started using Power BI. There are abundant resources for learning.

Use Cases & Future Directions

Rusch described the audience for the CPBI, which includes academic departments, university administration, librarian liaisons to the academic departments, and the library journal collection development committee. Academic departments can use the CPBI as the basis for accreditation reports and collection development discussions. Deans and others can learn more about the value of the library and how students use the library. Librarian liaisons can use the CPBI to support collection development, accreditation, and instruction. In addition, members of the library journal collection development committee refer to the CPBI often as they pursue their work. The CPBI continues to evolve as the CMT team receives feedback from users.

Use Cases

Rusch described four use cases for the CPBI: the accreditor visit, the department meeting, the package renewal, and the new program, followed by an outline of future directions. Earlier iterations of MNSU collection analysis reports contributed to previous accreditation successes, including positive comments specifically about the library and the journal collection, but the CPBI has improved how the library responds to accreditation needs. In the past year, the MNSU library's liaison to the Construction Management program, Heidi Southworth, was given two weeks' notice to prepare a Zoom presentation to accreditors from the American Council for Construction Education.

She developed the presentation including details about her library instruction, lists of relevant resources, and information drawn from the CPBI. For example, in the CPBI, she filtered on the subject building and construction. She found that the library supplies current access to forty-one of the forty-eight titles in the top quartile of the subject, so she was able to demonstrate adequate, or possibly superior, journal supply of high-quality journals for the program. Furthermore, when the accreditors visited campus and showed up at the library unannounced, Southworth was able to meet them and answer their impromptu questions utilizing the CPBI to look up relevant information in real-time. In the accreditation report, the reviewers praised Southworth and the library. They said they especially appreciated the journal data.

The MNSU library's liaison to the College of Business (COB), Lisa Baures, used the CPBI in a department meeting context to address COB collection development needs. Baures needed to balance a new journal package request against an ongoing collection review, whereby some journal packages were under consideration for possible cancellation. She led a college-wide discussion to convince the business faculty to support the cancellation of one journal package to offset the costs of the new request.

Baures used the CPBI in a live demonstration, starting with the CP Subject CSU report to show the depth of journal supply for the college and its programs. She filtered on several subjects in turn, such as accounting, business management, marketing, and so on. Next, she demonstrated the JL Subject CSU report to show the impacts of a journal package cancellation. She filtered on the subjects again and combined these filters with the package filter to demonstrate that the library would continue to provide recent coverage from an aggregator to nearly all the journals in the package under consideration for cancellation. Finally, she displayed the "CP YOP U" report. This report provides a column chart and a matrix, both based on COUNTER J4 data.[6] By filtering on the package, she showed that MNSU students and faculty mostly download articles from that package published more than one year ago, so the aggregator coverage would be sufficient to

meet ongoing needs, alongside interlibrary loan. In the past, Baures would have needed multiple static reports to make her case, but the CPBI allowed her to make her case in real-time, dynamically. The dynamism and aesthetic impact of the data viz certainly contributed to the impact of the presentation. According to Baures, the group of business faculty was "very interested in the . . . Collections Power BI."

Toward the end of the spring semester, the journal collection development committee discussed an upcoming package renewal. The package included some journal subscriptions that could be cancelled at the library's discretion, just like individual subscriptions. Members of the committee required only a few minutes to search these journals in the Journal Lookup report to decide whether to cancel any of these journals. One of the journals had a Southworth Ratio of .7, which indicates strong increasing usage, and the Article Downloads line chart showed how usage was tending upward over several years. The matrix showed that access to this journal was available from just one provider. Overall usage was high, and the journal was ranked in the first quartile for its subject. Clearly, this was a journal subscription to retain (see Figure 2). Other journals could be cancelled because they had Southworth Ratios below .5, combined with overlapping access from other providers, and lower overall usage. These searches required just a few minutes and left no doubt as to the appropriate decision.

Presenting the fourth and final use case, Rusch discussed a hypothetical problem based on previous experience. A few years before the CPBI was developed, Rusch was contacted by faculty to consult on the development of a new program in health informatics, specifically to address whether the library could provide sufficient journal supply to support the program. If the CPBI had been available then, he could simply have filtered on the subject health informatics and exported pertinent report elements to the documents he submitted in support of the new program. Similarly, any library liaison can use the CPBI to discover and explore the journals supporting any program, as well as any usage tendencies unique to MNSU.

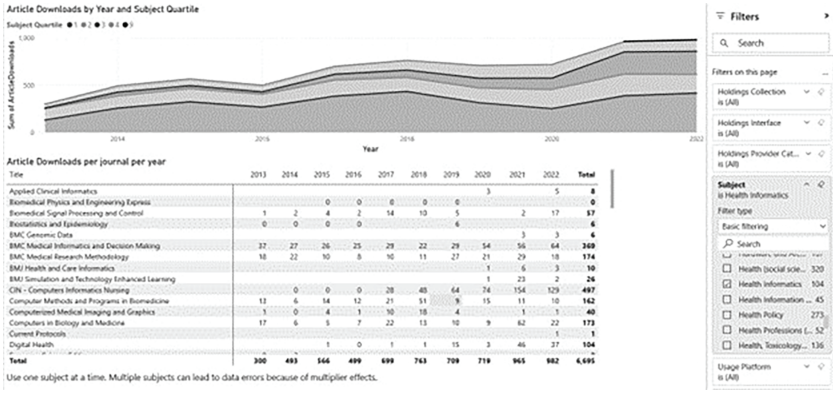


Figure 4. JL Subject U Trend

Rusch focused on two examples of reports he would use to investigate journal supply. For the first example, he showed a picture of the JL Subject QSU report filtered on Health Informatics and sorted on SciImago rank. He pointed out that users can see at a glance which journals have or lack current access. He can also see which publishers are important to the subject area based on both quality and usage. For the second example, he displayed a picture of the "JL Subject U Trend" report, which provides two data visualizations, a stacked area chart and a matrix. The chart displays aggregated article downloads data over the past nine years broken out by journal quartile. The matrix provides article downloads per journal for the past nine years. Filtered on health informatics, Rusch pointed out that usage of titles has almost doubled since that program began. This could provide strong evidence to support accreditation or for other program evaluation purposes (see Figure 4).

Future Directions

One of the drivers for developing the CPBI was to create a tool that could be used to inform campus-wide conversations about the impacts for specific programs of cancelling a "Big Deal." The team expects

to pursue these conversations over the next year or so. They will revise the CPBI as necessary. They also expect they will need to create simplified reports for some users, such as one-sheeters comprising elements exported from the CPBI. The team plans to develop a "cookbook" including recipes providing instructions for basic uses of the CPBI. These recipes could be focused on specific kinds of questions or contexts, such as accreditation. The team feels it is very important to engage university administration specifically and has established the goal of helping the Library & Learning Dean become an effective user of the CPBI. The Dean might have opportunities to present the CPBI in administrative meetings, just as Southworth and Baures presented to accreditors and faculty. The CPBI could also be developed to include more kinds of library data. A previous version included a report of all physical resource types in the library, such as books, but they imagine a future version could include data beyond collections.

Questions & Conclusion

Several questions focused on the data sources and data processing. These are best answered by reference to the MNSU Collection Analysis library guide. One member of the audience asked if the CPBI could integrate UnSub data.[4] Generally speaking, the CPBI can integrate any journal data. The CPBI and UnSub are quite different in purpose and scope. UnSub is limited to fewer data sources and is typically used to analyze individual journal packages. The CPBI includes (or can include) all the data and functionality of UnSub, but also much more, so the CMT team would not use UnSub. Instead, if there was anything particularly useful in UnSub, or any other third-party product, not currently possible with the CPBI, they would simply re-build that functionality within the CPBI.

The CMT team concluded the presentation by reminding the audience that interactive collection analysis and data visualization solutions like the CPBI would not be difficult to implement at other libraries. Any

library could process journal data as key lists and match lists. The data preparation does not require much time. These lists work perfectly as the basis for reports in Power BI – and Power BI itself is a cinch.

Contributor Notes

Nat Gustafson-Sundell is Associate Professor and Collections Librarian, Minnesota State University, Mankato, Minnesota.

Evan Rusch is Associate Professor and Reference/ Instruction Librarian, Minnesota State University, Mankato, Minnesota.

Pat Lienemann is Associate Professor and eAccess & Discovery Librarian, Minnesota State University, Mankato, Minnesota.

Jeff Rosamond is Technical Services Technician, Minnesota State University, Mankato, Minnesota.

Notes

1 "The SCImago Journal & Country Rank is a publicly available portal that includes the journals and country scientific indicators developed from the information contained in the Scopus® database (Elsevier B.V.). These indicators can be used to assess and analyze scientific domains." See the following link for more information, https://www.scimagojr.com/aboutus.php.
2 Minnesota State University, Mankato, "Collection Analysis," https://libguides.mnsu.edu/collection-analysis/
3 Nat Gustafson-Sundell, Kellian Clink, and Evan Rusch, "Combine Journal Data to Support Reference and Instruction," *Internet Reference Services Quarterly* 27, no. 3 (2023): 159–77, https://doi.org/10.1080/10875301.2023.2209573.
4 "The Unsub dashboard helps you reevaluate your [journal package] deal's value, and understand your cancellation options." See the following link for more information, https://unsub.org/.