Midwinter Online Conference 2023

Questions from Session 9: Automating a Data Pipeline for a Rapid Weeding Project Dashboard, by David Arredondo and Matt McDowall (University of Nebraska at Kearney)

Q1: How did you discover PowerBI? Did you evaluate any other BI solutions?

I was working on a different project last spring (2022) in which I was analyzing our University system's research output, specifically using Scopus as the primary data source. I knew I wanted to combine the data with Unpaywall data to see our OA output, and to combine APC list prices if available. My first instinct was to use Tableau Public for visualizing the data, and OpenRefine to combine and clean data. To get more than the Public version of Tableau was around $70/month, which I wanted so I could use some of the advanced features. I was aware of PowerBI but hadn't used it at that point, just from investigating all the different programs available in our Enterprise Microsoft account. I did not investigate any other programs once I found we had access, and I used O'Reilly for Higher Education as a resource to learn the program - there are good video-based books in that database.

Q2: Have you tried loading the data directly into Power BI using the Alma API?

We have not - this could be very useful and we haven't discussed it, so thank you for the idea!

Q3: could you please show the Github link again?

https://github.com/MattMcDowall/AnalyticsToSharePoint — Happy to accept suggestions/collaboration, btw, if you see ways to improve it

Q4: This was a for a weeding project. How did that project meet the objectives?

Our stacks was 235k volumes, and our target was to weed 20-25%, so in the 50k-60k range, which we knew we wouldn't achieve but wanted to get as much done as possible before the collection would be shifted to another floor of the library to begin renovation. In roughly 8 weeks, we were able to withdraw about 45k titles - it took every employee and student workers working around the clock. It was rough...Most titles we were able to box and send to BetterWorld Books, but many were in very poor condition and we had to rip covers and recycle.

Q5: Yes, just another BI tool considered?

We did not consider other BI tools - I'm not sure I'm aware of any so please let me know if there is one you are using!

Q6: What data determined something to be weeded?

This was another component of the project which Matt and I worked on and used PowerBI for. We ran an Analytics report with a robust amount of Bib and Item data for the Stacks collection location. We then split it into subject categories based on LC subclass, and Librarians were assigned to each section based on subject expertise and liaison assignment. Matt also ran a holdings report out of OCLC (how many other libraries in the US had holdings on a certain title), though I reviewed that separately because we got the data later on in the project. Librarians did a first pass, looking at circulation counts (we migrated Sierra data which included checkout and reshelving numbers, but they were combined), publication data, subject headings, series information, authors. After Librarians made first decisions on retention/withdrawal, they sent them out to academic departments for review. Some departments were very involved (History, English, Music), while others were fine with whatever decisions we made. Of course, some individuals were quite unhappy, though most were understanding of the situation. We initially wanted to do a shelf review after the spreadsheets were processed and circulation had flagged all the titles that were marked to withdraw, but we simply did not have enough time. When we pulled the titles, we scanned the barcodes with bluetooth barcode scanners which went to our catalogers to withdraw from the system at the end of each day, and then we loaded the OCNs to OCLC to batch delete holdings.

Q7: Did you consider another BI Tool vs. Power BI?

The only other tool I might have considered that I'm aware of is Tableau Public, but I made the switch to PowerBI in Spring 2022 and haven't looked back.

Q8: Who looks at the data/results? I mean outside of the library

For this project, no one outside the library was looking at the Dashboard - it was just used internally. If necessary though you can make the App public, or controlled access, or even embed them in Websites like a Libguide, which we hope to do with future reports using an automated pipeline. That might be more for department APR reports that we regularly get requests for (how much did we spend on books, databases, title counts, etc, for the discipline).

Q9: More of a comment than a question (I know! Sorry): if the goal was to show how close you were getting to the goal, did you consider including the "finisheing line? e.g. 68,578/goal, or your running total as a percentage?

This is a great idea that I didn't explore, but I'm sure you could do that in PowerBI - thanks for the idea! That would be especially helpful for certain projects (eg, we have to make cuts to our subscriptions this year and have a target number to hit).

Resources/insights shared by attendees in chat:

- GitHub for Python script: https://github.com/MattMcDowall/AnalyticsToSharePoint
- Link to join DASIG (Data Analysis and Services Interest Group): https://github.com/dasiglib
- We are actually doing this for PowerBI and when we export we are actually exporting them as CSV making them much easier to deal with - Using power automate