2017

MAPing Collection Use: Using Massive Analysis Projects for Collections Analysis

Galadriel Chilton
galadriel.chilton@yale.edu

Joelle Thomas
Trinity College, Hartford Connecticut, joelle.thomas@trincoll.edu

Alice Fairfield
University of Connecticut, alice.fairfield@uconn.edu

Arta Dobbs
University of Connecticut, dobbs@uchc.edu

Elisabeth Umpleby
University of Connecticut, elisabeth.umpleby@uconn.edu

See next page for additional authors

Authors
Galadriel Chilton, Joelle Thomas, Alice Fairfield, Arta Dobbs, Elisabeth Umpleby, and Dawn Cadogan

MAPing Collection Use:

Using Massive Analysis Projects for Collections Analysis

Assessing resources that are similar in size, scope, coverage, and typical user population. While this recipe was first used at an academic institution, it could be adapted for public, state, school, or special libraries.

Galadriel Chilton, Ivy Plus Libraries; Joelle Thomas, Trinity College; Alice Fairfield, University of Connecticut; Arta Dobbs, University of Connecticut Health Center; Elisabeth Umpleby, University of Connecticut School of Law; and Dawn Cadogan, Amherst College

NUTRITION INFORMATION

In spring 2014 at the University of Connecticut (UConn), a six-member working group completed a massive analysis of Scopus, Web of Science, and Google Scholar. The goal was to use many different quantitative data sources and qualitative data to create a comprehensive narrative of how the scholarly community used Scopus, Web of Science, and Google Scholar, why they were using each resource, and which resource features were most important.

DIETARY STANDARDS

ACRL *Standards for Libraries in Higher Education* (2011) Principle 4, Indicator 4.4; Principle 5, Indicator 5.1

COOKING TIME

Approximately 200 hours

COOKING TECHNIQUE

Analysis of multiple sources of qualitative and quantitative data to determine what each data source suggested about the use of Google Scholar, Scopus, and Web of Science by users. Then we compared and contrasted data to determine overall trends of resource preference and usage patterns.

INGREDIENTS

- A literature review for each resource
- COUNTER usage reports
- ILL requests
- EZ Proxy logins
- Coverage title lists
- Database A–Z list click-throughs
- Open URL referring source (e.g., SFX, Serials Solutions, etc.)
- User survey
- Environmental scan of access at peer organizations
- Platform functional comparison (e.g., 10 citation test)
- Usage reports for specific resource functionality

Ingredient Notes

For usage statistics and cost-per-use calculations, we used search and session reports as well as the cost for three calendar years. When it comes to ILL requests, we used reports from ILLiad for three years that showed the number of requests submitted by faculty or graduate students for citations from Scopus, Web of Science, Google Scholar, and PubMed. The EZ Proxy log yielded off-campus logins to Web of Science and Scopus by status and department. We downloaded coverage lists for Scopus and Web of Science, imported them into MS Access, and ran queries based on ISSN to compare coverage. Our home-grown database A-Z list enabled us to capture click-throughs from our database A-Z list. Open URL referring source reports showed the number of times an Open URL link was clicked from Web of Science, Scopus, or Google Scholar.

For qualitative data, subject librarians sent the following questions to department heads, department faculty, and graduate students in their subject areas:

- Do you regularly make use of either Scopus, Web of Science, or Google Scholar?
- Which resource do you use most often?
- In the resource most frequently used, how do you use the search functionality compared to the analysis features? (Please consider frequency of use, importance to you, etc.)

- If you use more than one of these tools, why do use more than one? What features do not overlap?

The environmental scan of peer organizations included a review of Scopus and Web of Science access via *US News and World Report*'s 2014 "Top 25 Public Universities" and UConn's Peer Institutions according to UConn's Office of Institutional Research. For the functionality test, we generated a list of ten citations representing a range of years and disciplines to search in each platform to compare functionality. Elsevier and Thomson Reuters provided non-COUNTER usage reports by usage type (e.g., analysis usage).

PREPARATION

Determine two or more resources that are similar in their scale and scope. Based on the resources being assessed, select the ingredients, collect data, and then compare and contrast results to reach a conclusion.

THE ASSESSMENT

Chefs include library staff who work with the primary audience that use the two resources being assessed (e.g., subject department). Chefs also include library staff with access to, and familiarity with, the ingredients used as well as having Excel and Access skills.

Before proceeding with the assessment, cooks should check with their library administration and, if an academic institution, their IRB office, to review policies and procedures around collecting and using information from the user community.

Next, develop your organization-specific ingredients list along with colleagues/departments (e.g., IT) that will need to help you retrieve data, and set timeframes for obtaining each data set. Next, determine which cook will be responsible for obtaining and analyzing each ingredient and writing the summary of findings for each. (See ingredient notes for ideas of how to analyze ingredients.) For quantitative data, we aimed to collect the same three consecutive years of data from each ingredient.

To prepare the report of findings, write an introduction describing the environment, circumstances, scale, and scope of the Massive Analysis Project.

For each data set, write a narrative summary describing how the data was collected, limitations of the data (e.g., EZ Proxy logs reveal usage off campus only), the results of the analysis for what the data reveals, and compare the data for each of the two resources. For example, how does the cost-per-session/search for one resource compare to the other?

Additionally, create a chart, graph, or other data visualization to accompany the narrative for each set of data (e.g., cost-per-search/per session, ILL requests originating from each resource, etc.).

Next, look for trends across ingredients: Where do comparisons of different ingredients show similar or disparate results?

Write a conclusion with recommendations and the pros and cons of each recommendation (e.g., keeping both resources, canceling both resources, keeping just one of the resources).

The final report should include an executive summary of methods and findings, an introduction, methods, a description and analysis of each ingredient and data source, limitations of the analysis, recommendations, a conclusion, and sources cited.

ALLERGY WARNING

Some ingredients may not be available for all resources being assessed. Also, data gathering and analysis will take a substantial amount of time. The Massive Analysis of Scopus, Web of Science, and Google Scholar took over 200 hours of staff time—including the group's seven meetings. (Most of our work was completed outside of meeting times.)

If sharing the results of the assessment in a public forum, be cautious of how cost/cost-per-use data is displayed. You may wish to scrub data points so that subscription costs cannot be calculated from information presented. At a minimum, check your library's policies and practices about publically sharing costs for e-resources, and check the resources' license agreements to ensure

that sharing data does not breech the organization's contract with the information provider.

CHEF'S NOTE

Chefs for this recipe should include representatives of library staff familiar with all aspects of resource usage. Our working group included: the head of e-resources, one subject librarian from Sciences and one from Social Sciences and Arts & Humanities, and librarians from UConn Health and UConn.

This Massive Analysis Project yielded a sixty-three-page report of findings. Having the massive amounts of data behind the group's recommendations was essential not only for decision-making but also to engage with, and respond to, our community. The methods used in the project are a recipe for collection assessment using qualitative and quantitative data to evaluate e-resource functionality and user preferences.

Future applications and variations would be less time intensive. Depending on the resource(s) evaluated, the recipe's ingredients could be scaled down from a multiple course meal to an entrée, side dish, or appetizer.