**Organization:** University of Rochester, River Campus Libraries  
**Primary mentor:** Sarah Pugachev, Director of Research Initiatives, River Campus Libraries, University of Rochester  
**Supporting mentors:** Sarah Siddiqui, STEM Librarian, Carlson Library, University of Rochester; and Allegra Tennis, Science and Engineering Outreach Librarian, Carlson Library

<table>
<thead>
<tr>
<th>Project Title</th>
<th>Using Data to make Impactful Collection Decisions</th>
</tr>
</thead>
</table>
| **Description** | In the River Campus Libraries Research Initiatives department, we have a focus on understanding the research being produced at the University of Rochester and how it is represented and visible in various databases including Web of Science, Scopus, and Dimensions.  
At the same time, like many of our colleagues across the nation, we are also faced with reducing our collections budget. We strive to incorporate data in order to help us make the best decisions for our researchers and students.  
In this project, the fellow will combine data about our collections with data from Scopus, to help us determine how we can make collections decisions that incorporate usage data as well as information about our communities publishing practice to ensure that our funds have the most positive impact possible.  
We have an initial start on this analysis from graduate students in our data science program focused on one publisher, Wiley, but we would like to expand this analysis to include a broader look at our collections.  
This project can be adjusted based the interest of the fellows and they desired learning outcomes. |
| Problems/Research Questions | These key research questions are related to better understanding our collection usage and publishing practices.  
- What journals of those that we subscribe to have the most usage?  
- What is the cost per usage for different journals?  
- What journals do UR faculty publish in frequently?  
- What journals do UR faculty cite most frequently?  
- What is the relationship (if any) to where faculty are publishing and what they are reading?  
- Additionally, addressing these questions (if time allows) will allow us to better understand how publishing and research practices can affect the visibility of research being conducted at UR.  
- Which of the journals that we publish in are the most impactful (according to available citation metrics)? Does this correspond with usage or cost?  
- Who are UR authors collaborating with? Where do those collaborations get published? |
| Techniques | Data visualization, bibliometric analysis, network analysis, time-series analysis, regression and prediction models, or other relevant techniques as identified by the fellow |
| Tools/Languages used | Python preferred or open to other languages depending on fellow’s skillset; Tableau, UnSub, Web of Science or Scopus |
| Data | There are two different types of data associated with this project. The data will be provided as spreadsheets or training will be given on a system to allow the fellow to pull relevant data themselves.  
**Collections Data**  
Description: Spreadsheets of usage statistics from the UnSub and for different journals from our ILS (Alma); these will contain journal name, issn, usage, and cost information along with other metadata. This will be supplemented by usage data provided by our Acquisitions department, which is available by year.  
Data Type: spreadsheets (CSV, Excel) |
### Publication data

description: Publications of UR faculty is available through two databases (Scopus and Web of Science). These systems allow access to information including article titles, journals year, citations, topics, and more. A sample set of data will be provided and the fellow will receive training on access additional data from these systems as needed.

data type: spreadsheets (CSV, Excel)

### Outcome

reproducible code and visualizations that we can use annually to assess our collection decisions based on usage, cost, and publications data.

### Milestone Timeline

Below is an estimated timeline with milestones. These are subject to change depending on how the project progresses.

- **June-July**: Introduction to project, research questions, and basics of bibliometrics. Understand data and perform data cleaning if necessary. Understand and receive training on systems used to access additional data.

- **August- September**: Determine and group journals based on usage data and cost. Determine most commonly published in journals by UR faculty and most impactful considering discipline.

- **September- October**: Compare, analyze, and visualize journals based on usage, cost, and publication practices.

- **November**: Make recommendations for collections decisions and document the process used to analyze data to reach those recommendations.

- **November - December**: If time permits, consider collaborations within the publications and if there is a correlation with usage or impact.

### References


Martín-Martín, Alberto, Enrique Orduna-Malea, Mike Thelwall, and Emilio Delgado López-Cózar. 2018. “Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories.”
|---|
  https://editorresources.taylorandfrancis.com/understanding-research-metrics/. |