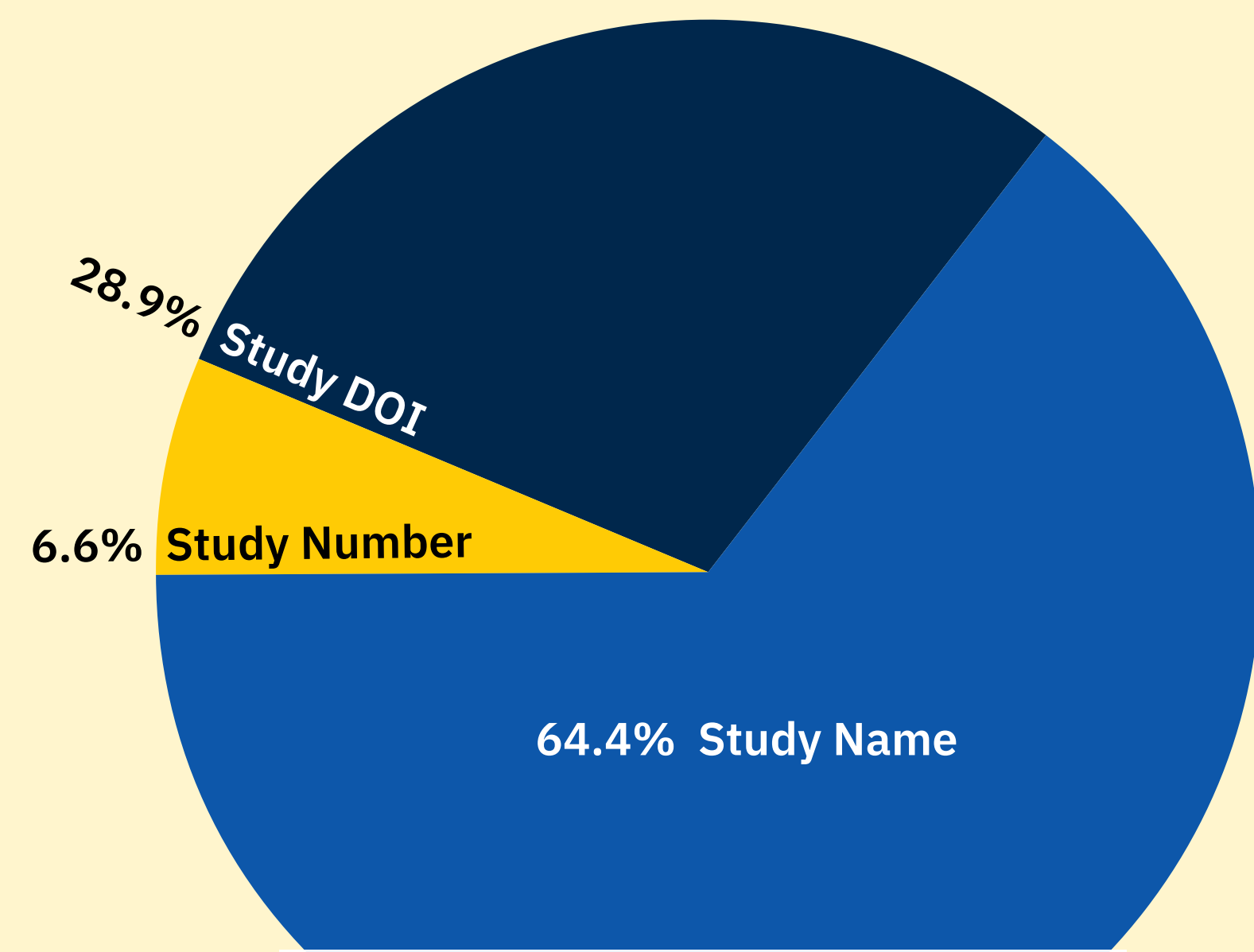
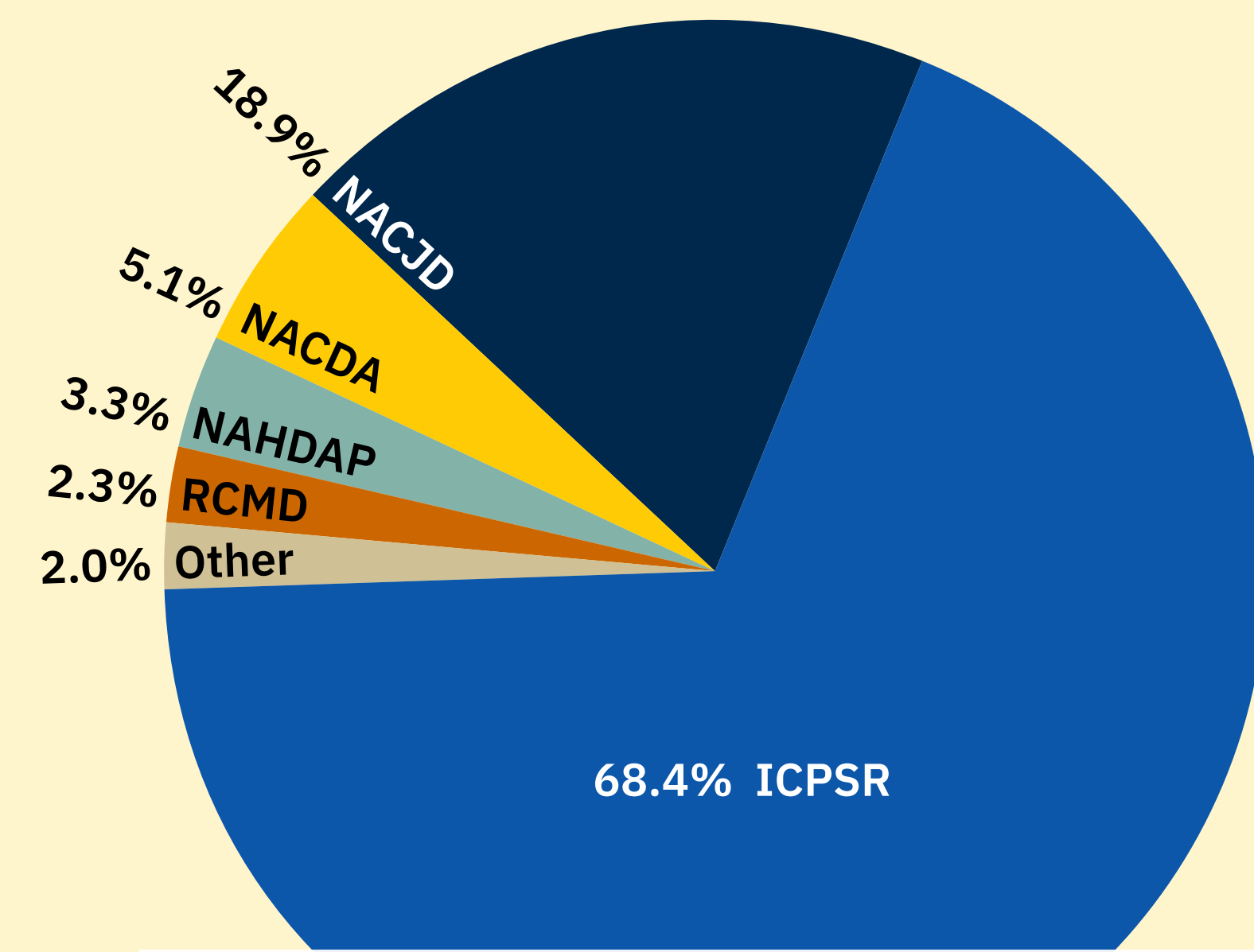


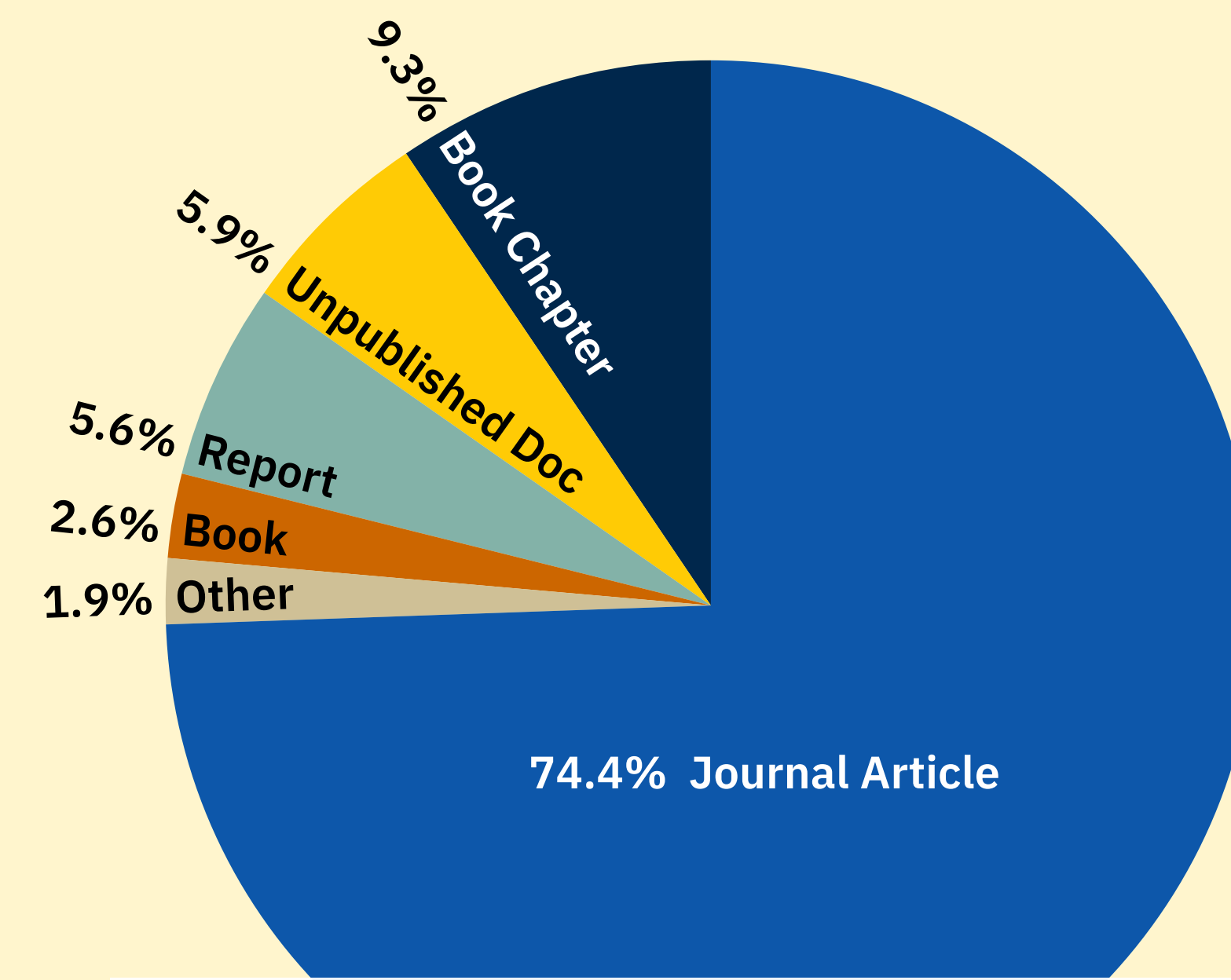
Results 2,532 records added to the Bibliography from Dimensions as of 3/22/24



Additions by **Query** from Dimensions



Additions by **ICPSR Archive** from Dimensions



Additions by **Publication Type** from Dimensions

Methodology

The Bibliography of Data-related Literature includes more than 113k items.¹ ICPSR has 11,763 studies.² The bulk of the Bibliography has been collected very manually with staff literally combing through literature to find instances of data analysis. A former postdoc worked with the Bibliography team to create a program using Dimensions, a large and comprehensive linked database, to programmatically find publications. The following approach was built upon this original program.

Step 1 Collect Information from ICPSR's Backend Database

Access the backend database to retrieve all Bibliography entries and study metadata.

Step 1b - Select Version: Specify which version of the bibliography they want to search (e.g., older version or most recent).

Step 1c - Add DOIs: Add DOIs of Bibliography publications to a dataframe.

Step 2 Determine Scope

Specify whether to test the program using the first ten studies, a specific set of ICPSR archive's studies, or search for all studies.

Step 2b - Add ICPSR Study DOIs: Based on the selection, add study DOIs, names, and unique numbers to a new dataframe.

Step 3 Search Dimensions

Utilize the Dimensions API to search for publications that match any of the criteria from list 1 AND list 2

Step 3b: Add metadata of matching publications, including DOIs, to another dataframe.

List 1	List 2
Study DOI	"ICPSR"
Study Number	"Interuniversity Consortium for Political and Social Research"
Study Name	"Inter-university Consortium for Political and Social Research"

Step 4 Compare and Filter

Compare the bibliographic publications dataframe with the Dimensions publications dataframe. Ignore duplicates. Identify DOIs not found in the bibliography and input into a CSV list along with metadata.

Step 4b: Study DOIs, Study name, and Study number are added to the CSV on the matching publication with a key indicating the type of query used (study name, study number, or study DOI).

All Steps Progress Display

Print progress updates in the terminal as the search progresses (e.g., "Searching for publications matching study DOIs...").

Improvements to the original code

- Enabled searches across multiple archives through allowing user to choose more than one archive to search
- Added ability to exclude specific ICPSR archives from a search
- Facilitated Great Lakes access to each staff member so they could run searches using High Performance Computing
- Wrote detailed instructions for the Bibliography staff to perform searches on their own
- Instituted "date_inserted" field in Dimensions query, reducing the amount of time spent reevaluating bad hits
- Added timeout protection to the code, ensuring the program didn't error out

Challenges

- Amount of data to sift through and determine usage
- The size of data to compare is large and only possible on external HPC machines
- Data citation and naming practices continue to make it difficult to assess data usage
- Staff must divide their time based on who is funding the work, making it difficult to find time to evaluate Dimensions results
- Determining data use is time consuming in the best conditions

Acknowledgements

Thank you to Sara Lafia for the original code. View the original code here:

<https://github.com/ICPSR/search-dimensions/tree/main> 🔍

Future Plans

- Regular API pulls from Dimensions for Bibliography staff
- Expanded capability to find publications based on more fuzzy matches (not relying on DOIs or other persistent identifiers)
- Ability to search more flexibly using one or more study numbers specified by the user
- Other APIs — this is difficult due to very few generalist databases allowing API searching

ICPSR is part of the Institute for Social Research at the University of Michigan.

Endnotes: 1) though only 112,781 are visible to the public 2) number does not include openICPSR as the Bibliography team does not collect for those studies at this time