Sweeten the Search
Discover Data for Reuse with a Tool That Links Publications to the Underlying Data

Elizabeth Moss, Librarian, Inter-university Consortium for Political and Social Research (ICPSR), Institute for Social Research, University of Michigan, eammoss@umich.edu

NUTRITION INFORMATION
It can be challenging to find shared social and behavioral health data with the appropriate ingredients for reuse in class assignments or academic research. Help sweeten the data discovery process with a tool that is especially appealing for students who are hesitant about using quantitative data. In a three-part activity, students start by creating a user story to assess needs and determine the best terms for an effective search. Then they take an active tour of a freely available collection of scholarly literature linked to the wide array of social and behavioral research data available for reuse at a large research data archive. While on the tour, students perform searches to discover relevant data. After the tour, they perform a final self-assessment of their data discovery process.

TARGET AUDIENCE
Undergraduate and graduate students

LEARNING OBJECTIVES
Students will
- construct a user story to inform a search strategy
- tour and conduct searches in a research data archive's data-related publications database to find and recognize relevant content when evaluating data for reuse
- assess their data discovery process

COOKING TIME
45 minutes

DIETARY GUIDELINES
ACRL's *Framework for Information Literacy for Higher Education* frames: Research as Inquiry; Searching as Strategic Exploration

INGREDIENTS
- Web browser access for students and librarians
- 1 copy per student of User Story Template and Examples (appendix A)
- 1 copy per student of Tour of the ICPSR Bibliography of Data-Related Literature (appendix B)
- 1 copy per student of Discovery Process Self-Assessment Template and Example (appendix C)

PREPARATION
Before holding the class
- Adapt the User Story Template (appendix A) by customizing the examples to align with the likely subject area of the class members.
- Create an example user story to use when instructing the students during Activity 1.
- Go through the Tour document (appendix B) yourself so you are familiar with a search strategy and outcome you want to use as an example for the class during Activity 2.
- Prepare ahead for Activity 3 by anticipating how you would evaluate your example discovery experience when demonstrating how to fill out the short, six-step Self-Assessment document (appendix C).
- Provide an online location for students to access the three documents (appendices A, B, and C) that they will use in the class, or print them out and provide one copy of each of the three documents to each student.

INSTRUCTIONS
Explain to the class that they will be learning how to find research data of interest to them by searching a large database of publications that have been linked to the data underlying the findings, called the "ICPSR Bibliography of Data-Related Literature." The data being analyzed in the publications are located at

ICPSR, an archive of shared social and behavioral research data. Stress that before you begin searching any database, it is best to get clarity about what you are searching for, and why. This leads you to Activity 1.

Activity 1
- Ask students to open a browser window or tab to access the user story examples provided via the link in the document.
- Once they have reviewed the examples, guide class members through the User Story Template in appendix A by using the example you prepared.
- Then ask them to create their own user story statements and to note specific data needs and search terms to use.
- Encourage them to think about alternative or synonymous terms to use when searching the database.
- When class members have finished, move to Activity 2.

Activity 2
- Together with the students, read the introductory paragraph of the Tour document in appendix B.
- Important: Point out to the students the information in this recipe's Chef's Notes.
- Together with the students, open a new browser window or tab and navigate to the links provided in the document in Step 1 and Step 2.
- Answer any questions they may have about the Search Tips (from Step 2).
- With the class, go to Step 3 in the Tour document and demonstrate entering the search terms you prepared. Show

how you can try different terms if the first combination is unsuccessful.
- Briefly demonstrate sorting and filtering in Step 4.
- For Step 5, demonstrate choosing a search result that interests you based on the needs identified in your user story. Show how to click out to the full text of the publication, as well as how to link to the study or studies associated with that publication. Tips:
  – *When in a publication:* Point out why any aspect discussed catches your eye; for example, the authors discussed an interesting approach they used to analyze the data; the authors pointed out specific variables that you would want to analyze; the authors mention that they used the public-use or restricted-use version of the data for a particular reason that is relevant to you; the authors were able to use one data set in combination with another to achieve an interesting finding; unique aspects of sample population match your needs; the authors discuss in-depth the methodology used to collect the data that makes you want to access the data for your own use; you are reading a report with summary statistics—and that is what you want to paste in your term paper and cite, and so on.
  – *When in the study description:* Point out what may be of interest to you; for example, the sample size, population, or location is appropriate for your needs; you want to download the codebook to become more

familiar with the data; the data are available for online analysis; the data files you need are publicly available; there are other data-related publications associated with the study that you want to look at more closely, and so on.
- Once you have finished your demonstration, invite the students to type their terms in the search box and begin the process that you just demonstrated, as outlined in Steps 3–5 in the Tour document.

Activity 3
- Once students have completed the Tour, at least ten minutes prior to the end of the class, tell the students it is time to document their experience by filling out the Self-Assessment document (appendix C).
- Using the six assessment aspects listed in the document, briefly state out loud your self-assessment for the example discovery experience you provided during the class.
- Ask the class to take five minutes to fill out their own self-assessment document based on their own discovery experiences.
- End the class by encouraging the students to continue exploring this resource for current or future data needs.

REVIEWS/ASSESSMENT STRATEGY
See appendix C, "Discovery Process Self-Assessment Template and Example"

CHEF'S NOTES

- Be sure students understand ICPSR's definition of a study. At ICPSR, a study is a collection of one or more data files and the documentation needed to understand how to use those files. Each study has its own description page on the ICPSR website.
- Be sure students understand that they can search ICPSR's study catalog directly for data. Point out that the Bibliography of Data-Related Literature is an additional resource allowing them to find data via the publications using the studies in the ICPSR catalog. That resource is the focus of this class.
- The ICPSR Bibliography does not index the full text of publications. While searching the bibliography yields fewer results than searching the study catalog, those results are very relevant. If a word is important enough to go into the citation, then the publication often will have significant content using that word or concept.

ADDITIONAL RESOURCES

Association of College and Research Libraries. *Framework for Information Literacy for Higher Education*. Chicago: Association of College and Research Libraries, 2016. https://www.ala.org/acrl/standards/ilframework.

ICPSR. "ICPSR Bibliography of Data-Related Literature." 2021. https://www.icpsr.umich.edu/web/pages/ICPSR/citations/.

ICPSR. "User Stories." 2021. https://www.icpsr.umich.edu/web/pages/ICPSR/citations/user-stories.html.

Appendix A: User Story Template and Examples

Create a user story to clarify search goals and data needs.

Start by examining the existing user stories for ICPSR Bibliography users, found on the ICPSR website:

> https://www.icpsr.umich.edu/web/pages/ICPSR/citations/user-stories.html

Create a user story, incorporating this statement to state your goals:

> **As a <user role>, I want <goal> so that <benefit>.**

> **Example:**

> **As a** senior in a capstone course, **I want** to find data I can use to write a paper **so that** I can generate statistics online and use them as part of the assigned work.

In addition, note specific data needs to consider when examining your search results.

> **Example:**

> I need data containing variables about high school students and attitudes about race and equality.

Based on the identified goals and data needs, list likely search terms to use when conducting the database search.

> **Example:**

> "race relations" "high school students"

Appendix B: Tour of the ICPSR Bibliography of Data-Related Literature

With your user story and search terms, you are ready to find social science or behavioral health research data that will meet your goals. You will tour the ICPSR Bibliography of Data-Related Literature and search for scholarly works of interest as you go. Those works are all linked directly to the full text of the publications (where possible), *and* they are linked to the data used in the publication. Those data are described and archived at the Inter-university Consortium for Political and Social Research (ICPSR), a large repository of curated, digitized social and behavioral health science data archived for people like you to access and reuse.

Step 1: Start your tour at the ICPSR Bibliography's search portal: https://www.icpsr.umich.edu/web/pages/ICPSR/citations/

Step 2: Review the search tips located just below the main search box.

Step 3: Enter in the search box the term or terms defined in your user story.

Example: "race relations" "high school"

- Look for search results to be returned in the tab called "Data-related Publications."
- This search yields no results in the "Data-related Publications" tab. A modified (in this case, broader) query, e.g., *race* and *"high school"*, will yield many publications related to many data collections, which we will refer to as studies. ICPSR studies contain one or more data files together with documentation files needed to understand how to use the raw data. Each study can be found on a study home page, which contains links to these files, along with an extensive description of the study and download links.

Step 4: Once you have a set of search results, you can change how the results are sorted, and filter by:

- *Publication year:* This works well for students interested in only the most recent publications.
- *Publication type:* There are 11 different types, and you may want to see only journal articles. If you do not intend to download

data and you want to use the database to find publications containing summarized statistics that you can quote, you may want to limit your results to reports.

- *Journal title:* Specific journals are seen as significant in some fields of research, so you can limit your search results to see only what is in the database published in a particular journal.
- *Author:* You may want to see publications by a particular author whose work may be connected to available data. Keep in mind that the names in the database are not authority controlled, so they may appear in slightly varying forms, e.g., Smith, John A., could be the same as Smith, John.
- *Study:* Try clicking "view all" in this filter. It gives you a consolidated list of all the studies in the ICPSR collection that are associated with publications that are retrieved by your query.

Step 5: Now that you have sorted and filtered your results, you have reached a point in the tour where you can explore any search results of interest. You can choose to

- link to the full text, where possible, and read one or more publications to learn how the authors made use of the data
- select one or more citations to export into your bibliographic software
- click on the link to the underlying study or studies used in any publication to access the study's description page to
  – explore other publications listed in the Data-related Publications tab associated with that study so you can read more ways the data have been used.
  – read the study's description for information, e.g., to see if its data files are available for online analysis, to examine its variables, to read the methodology used to collect the data, to see the sample used, to read its online codebook, or to learn if there are restricted files in the study
  – click on the series link (it will display at the top of the study page if the study is part of a larger grouping of studies) to find more publications associated with the studies in the series

Appendix C: Discovery Process Self-Assessment Template and Example

Once your searching is done, complete a self-assessment and briefly document your

1. *Original goal:* (Restate your original user story statement.)
2. *Search terms:* (What terms worked?)
3. *Challenges:* (Were there any stumbling blocks?)
4. *Results:* (What goal was reached?)
5. *Surprises:* (Did you find something unexpected during the process?)
6. *Next steps:* (What direction can you pursue in future explorations of the database?)

Example

1. *Original goal:* As a senior in a capstone course, I wanted to find data I could use to write a paper so that I could generate statistics online and use them in the paper as part of the assigned work.
2. *Search terms:* race and "high school"
3. *Challenges:* I could not find any publications using my chosen terms, so I modified "race relations" to race so they were not so specific.
4. *Results:* I found a very relevant 2020 article in a peer-reviewed journal linked to a data set that I can analyze online. The article clearly used variables in the data that captured whether a respondent will be enrolled in college, as well as parental attitudes about college, and other race/ethnicity variables captured in the data. So I have confirmation that I can get the data I need to create statistics for my paper.
5. *Surprises:* I thought I would have to know more about a statistical analysis package to be able to generate statistics for my project. But I read the data description, which tells me I can use online analysis without having to know SPSS, SAS, or R!
6. *Next steps:* The paper I read gave me some great ideas for analysis, so not only can I use this data set for my assignment this weekend, using online analysis, but I also might want to try something more complex with the data in the future as I become more proficient in stats software.