

Continuous Curation of a Digital Bibliography Showcasing Criminal Justice Data Use

Sarah Burchart, Inter-university Consortium for Political and Social Research (ICPSR) • sdonetti@umich.edu





Since beginning work on the Bibliography for ICPSR's National Archive of Criminal Justice Data (NACJD) in January 2018, I have developed a routine, continuous methodology for finding and adding publications that use or analyze ICPSR data on a daily basis, resulting in adding over 4,000 new citations to the Bibliography.

Curating this data-related digital bibliography is a balance of a librarian's collection development and query construction plus automation to build on that human judgment and repeat it to continue finding new related publications.

- related publications
- Carefully study metadata for the dataset, (picking out contextual clues) in dataset queries, since the vast majority of
- Use appropriate search queries to probe scholarly databases, government websites, and other relevant resources (Examples: [11219 AND "sexual assault kits"]; ["sexual assault kit" AND "Detroit" OR "untested"])
- (Examine search results) to find articles using

- Select a dataset to initiate a search for
- metadata that would make good search publications using/analyzing ICPSR data don't formally cite it with the dataset's DOI
- or analyzing ICPSR datasets

2. Adding

- Enter new citations for publications found in a search that are not already included in the Bibliography
- Enable linked database to display related publications with data collections

- If a search query provides productive results, create an email alert within the scholarly database to receive updates when new publications with matching keywords appear
- Check NACJD's Bibliography email account on a daily basis for query alerts with new results matching collection criteria. Running a query happens continuously and automatically – I only have to examine the results brought back

Picking contextual clues in dataset metadata Study Design 2

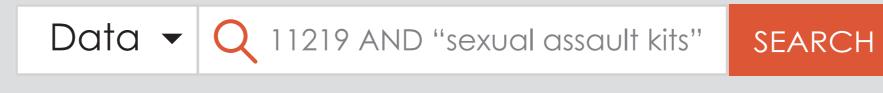
2) The second step involved investigating why so many SAKs were un-submitted. The research team questions regarding the utility of SAK testing under different case circumstand

Using search queries

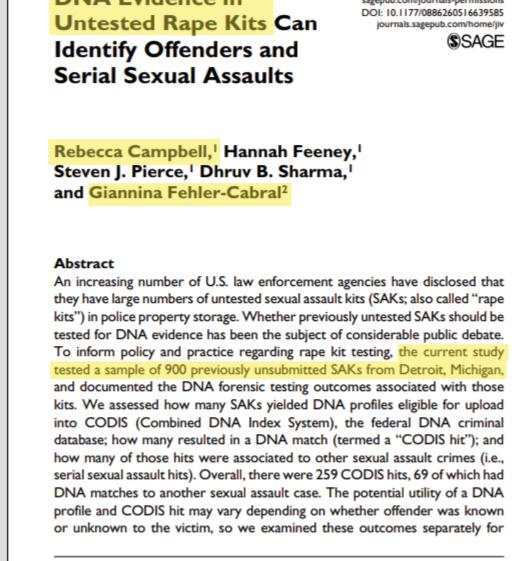
Tested at Last: How

¹Michigan State University, East Lansing, MI, USA

²Harder + Company Community Research, Los Angeles, CA, USA



Examining search results



the vast majority of which had never been tested for DNA (see Campbe

CODIS associations to other sexual assault cases. We examined these out-

comes for both stranger- and non-stranger-perpetrated sexual assaults to

lore the potential utility of DNA evidence and CODIS hits depending or

Bybee, & Fehler-Cabral, 2012; Campbell, Patterson, & Bybee, 2012 we defined stranger-perpetrated sexual assaults as those in which the victi

Journal of Interpersonal Violence 33(24)

ICPSR'S BIBLIOGRAPHY OF DATA-RELATED LITERATURE

To track the use of the data that ICPSR archives to facilitate greater understanding of its applications, the Bibliography connects datasets to citations of related journal articles, scholarly theses, and other publications using or analyzing the data.

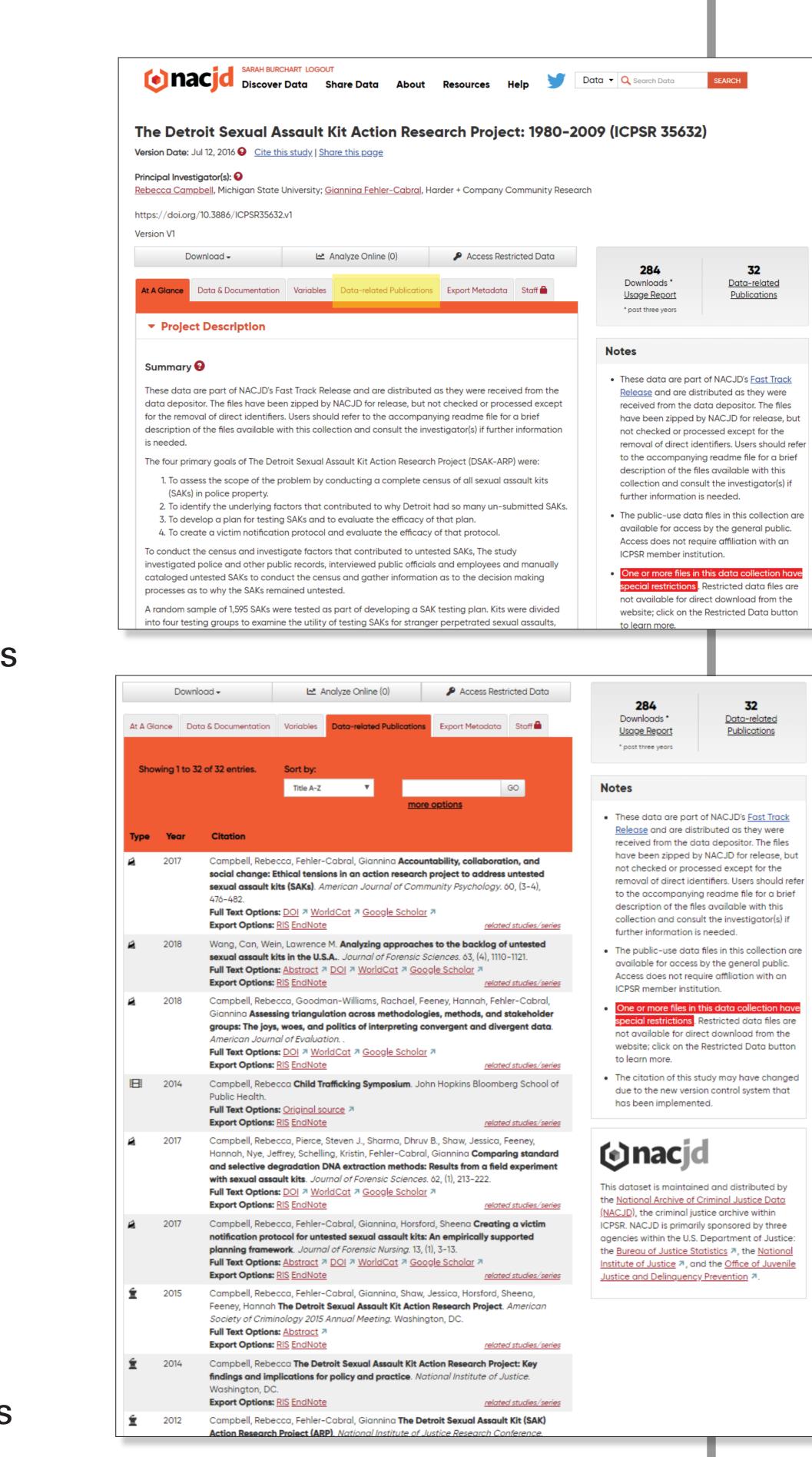
COLLECTION CRITERIA

To determine whether or not a publication should be included in the Bibliography, ICPSR librarians ask the following questions – does the publication:

- Analyze data archived by ICPSR?
- Discuss/critique/extend previous analyses of data archived by ICPSR in a way that points to the original raw data?
- Discuss/critique the data source or its design/methodology in depth?
- Compile new charts quoting the data?

RELEVANT LIBRARY SCIENCE COMPETENCIES

- Familiarity with ICPSR dataset metadata to make development decisions
- Balancing awareness of user requests, usage statistics, and relevance to current events when selecting a dataset for focused searching
- Regular examination of current subject-specific scholarly literature
- Construction of effective search queries based on the metadata and how data are referenced in scholarship
- Understanding of various academic database platforms and their respective advantages or disadvantages



CONTINUING QUESTIONS

- When ICPSR holds the data for over 2,700 criminal justice studies, how should I prioritize which studies receive the effort of focused searches?
- What is the best way to track my history of dataset searches and effective query alerts in an understandable and replicable way for other ICPSR librarians? (ex. Google Sheets document? Microsoft Access? Airtable?)
- How can future use of publisher APIs and machine learning widen the scope and scale of this work and increase automation without sacrificing the precision human judgment provides?

Fast Facts about the National Archive of Criminal **Justice Data:**

- Established in 1978
- Data from over 2,700 curated studies or statistical data series
- Sponsored by the Bureau of Justice Statistics, National Institute of Justice, and Office of Juvenile Justice and Delinquency Prevention
- Located within the larger Inter-university Consortium for Political and Social Research, a social science data archive at the University of Michigan

Fast Facts about the Bibliography of **Data-Related Literature:**

- Developed in 2000 with the assistance of a National Science Foundation grant
- Over 80,000 citations as of 2019 (at least 16,000 of which are connected to NACJD data)
- Shows how data holdings have been used and analyzed in research publications