An Introduction to CJARS:

A new data platform for integrated criminal justice research

Michael Mueller-Smith
University of Michigan

March 5\textsuperscript{th} and 9\textsuperscript{th}, 2020
Webinar topics

1. What is CJARS? Where to find detailed documentation?

2. Similarities and differences of CJARS to other statistical reporting initiatives on U.S. criminal justice

3. Accessing CJARS and research opportunities

4. Resources to support early-stage researchers
Several key resources will be introduced:

Data Documentation

Benchmarking Report

Proposal Guide

Fellowship Solicitation
- Live Q & A panel following presentation

- Submit questions via the “Q & A” tab in your Zoom meeting window

- FAQ to be posted to CJARS website in coming weeks
Defining the Criminal Justice Administrative Records System (CJARS)
What is CJARS?

CJARS, founded in 2016, is a joint data infrastructure project between University of Michigan and the U.S. Census Bureau

Building a novel data platform to modernize research and statistical reporting on the U.S. criminal justice system, including:

- Event-level criminal justice data with nationwide scope
- Tracking across key milestones in the justice system
- Capacity to link with individual-level survey and administrative data at the U.S. Census Bureau

Financial support from the National Science Foundation, the Bill and Melinda Gates Foundation, the Laura and John Arnold Foundation, the University of Michigan, and the U.S. Census Bureau
CJARS at a glance

Data Providers
- Federal Agencies
- State Agencies
- Local Agencies

U-M CJARS Repository
- U-M CJARS Secure Data Enclave Server
  - Administrative Data
  - Roster File
  - Physical Archive

U.S. Census Bureau
- Person Identification Validation Servers
  - Data Intake
  - Anonymized Research Servers
  - Disclosure Review
  - Project Approval

External Researchers
- External research request
- Non-CJARS external publication

CJARS statistical reports
Justice-involved individuals (JII)

CJARS data is oriented around the concept of the *justice-involved individual*
  - Does not presuppose guilt or innocence.

JII population includes:

- **Arrestees**
- **Criminal Defendants**
- **Inmates**
- **Probationers/Parolees**
Why do we need CJARS?

Key contributions over existing resources:

1. Longitudinal, multi-jurisdictional data is collected, harmonized, and linked to track individuals across space and time

2. Evolution of a criminal episode is traced through the justice system

3. Built for integration with socio-economic survey and administrative data held by the Census Bureau

4. Secure, responsible research access options without gatekeeping
What CJARS is not

Examples of groups and events that are out of scope:

- Minors involved in the juvenile justice system
- Reported crimes that do not result in an arrest
- Career progression of law enforcement, prosecutors, judges, or correctional officers
- Other events that might elevate criminal risk or be impacted by involvement in the criminal justice system (e.g. housing evictions, child welfare investigations, child support obligations, etc.)
CJARS is well documented
CJARS data documentation

Contents include:
- Project description and data collection methods
- Data holdings and structure
- Linkage process and other automated algorithms
- Variable codebook and descriptive statistics

Link: https://cjars.isr.umich.edu/data-documentation-download/
Extensive data holdings to date:

- 2+ billion lines of raw data
- 150+ million criminal justice events
- 30+ million unique individuals
- Statewide coverage in 1+ domains for 43% of U.S. population
Data documentation: relational structure

1 Jll roster and 5 event-specific data tables, with 2 ways to internally link:

- `cjars_id` to pool all events for one individual
- Event identifiers (`arr_id`, `adj_id`, etc) to follow an individual criminal episode through the justice system
Data documentation: detailed timelines

Historical coverage varies across geography and procedural domain:

Example timeline coverage from data documentation

Differentiates between partial and statewide coverage, snapshot data, and data collection efforts that are currently underway.
Data documentation: schema

National data schema balances between:
- Complexity to support nuanced research questions
- Consistency across jurisdictions to reduce barriers to working with data

Finite resources also limit the level of detail extracted from source files
- Data collection vs. data processing vs. data analysis

Where possible, we preserve source values (e.g. offense descriptions, sentencing fields) in addition to providing users with harmonized variables to support a flexible range of uses
Data harmonization innovations

Person-level probabilistic record linkage model

- Cutting edge machine learning model trained on fingerprint-validated identifiers
- See Mueller-Smith and Gross (2020) “Modernizing Person-Level Entity Resolution with Biometrically Linked Records”
- No biometric identifiers are transferred from the University of Michigan to the U.S. Census Bureau

Text-based Offense Classification (TOC) Tool

- Measures for Justice collaboration
- Translate over 4 million unique offense descriptions to 271 standard offense codes
- Public web-based interface to be launched later this year
Data documentation provides a description, summary statistics, and data visualization for every harmonized variable in the CJARS data platform.

<table>
<thead>
<tr>
<th>arr_arr_dt_yyyy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Table Format</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>Count</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All records</td>
<td>32,504,351</td>
<td>100.0</td>
</tr>
<tr>
<td>Valid records</td>
<td>20,402,688</td>
<td>62.8</td>
</tr>
<tr>
<td>Invalid values</td>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>Missing values</td>
<td>12,101,661</td>
<td>37.2</td>
</tr>
</tbody>
</table>
Data documentation provides a description, summary statistics, and data visualization for every harmonized variable in the CJARS data platform.

<table>
<thead>
<tr>
<th>Set</th>
<th>Count</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All records</td>
<td>32,504,351</td>
<td>100.0</td>
</tr>
<tr>
<td>Valid records</td>
<td>20,402,688</td>
<td>62.8</td>
</tr>
<tr>
<td>Invalid values</td>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>Missing values</td>
<td>12,101,661</td>
<td>37.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2,002.2</td>
</tr>
<tr>
<td>Minimum</td>
<td>1,901</td>
</tr>
<tr>
<td>25th percentile</td>
<td>1,996</td>
</tr>
<tr>
<td>Median</td>
<td>2,005</td>
</tr>
<tr>
<td>75th percentile</td>
<td>2,011</td>
</tr>
<tr>
<td>Maximum</td>
<td>2,020</td>
</tr>
</tbody>
</table>
Data documentation provides a description, summary statistics, and data visualization for every harmonized variable in the CJARS data platform.
Data documentation: variable codebook

Data documentation provides a description, summary statistics, and data visualization for every harmonized variable in the CJARS data platform.
CJARS has been validated against official federal statistical series
Why benchmarking is critical

Comparative strengths of CJARS come with potential vulnerabilities:
- Enterprising data collection (i.e. scraping, DUA’s, research agreements, and donations)
  ⇒ Inconsistent data structures and value codes
- Probabilistic matching algorithms in the absence of unique identifiers, and,
- Machine learning-augmented harmonization strategies

Need to validate the CJARS data product against existing statistical series
Assess reproducibility of official Bureau of Justice Statistics published series using CJARS micro-data, including:

- Uniform Crime Report (UCR)
- State Court Processing Statistics (SCPS)
- National Prisons Statistics Program (NPS)
- National Corrections Reporting Program (NCRP)
- Annual Probation and Annual Parole Surveys

Link: https://cjars.isr.umich.edu/benchmarking-report-download/
Examples exercises include:

- Caseload composition (e.g., demographic composition of felony defendants)
- Event counts (e.g., yearly entries onto parole)
- Caseload populations (e.g., yearly inmate population)

State or sub-state results where possible to align with our level of data acquisition and processing

Note: Estimates created using UM CJARS repository. See Papp and Mueller-Smith (2021).
Examples exercises include:

- Caseload composition (e.g., demographic composition of felony defendants)
- Event counts (e.g., yearly entries onto parole)
- Caseload populations (e.g., yearly inmate population)

State or sub-state results where possible to align with our level of data acquisition and processing

Prison Entries in Pennsylvania

CJARS can be accessed through the Federal Statistical Research Data Centers
Working with CJARS data

CJARS micro-data is currently available for external request to qualified researchers on approved projects through the FSRDC network
  – No CJARS project charges but FSRDC sites may have access fees

Public synthetic data product, which reproduces the underlying variation in CJARS without compromising individual privacy or confidentiality in development
  – Expected release in 2022q3

Public data portal for aggregate statistical information about caseloads and socioeconomic outcomes under development
  – Launch date is to be determined
Title 13 proposal guide

FSRDC research is authorized under United States Code, Title 13
- Meeting Title 13 requirements can be an opaque process for those without prior experience

CJARS Title 13 Proposal Guide will help facilitate successful CJARS-based research proposals. It covers:
- Requesting CJARS data access through the U.S. Census Bureau’s FSRDC system
- Proposal development and requirements
- Guidance on how to meet Title 13 statutory requirements
- Where to find information on other linkable datasets held in the FSRDC system
- Bringing in your own data to the FSRDC system
- Review timelines and what to expect after proposal submission

Link: https://cjars.isr.umich.edu/proposal-guide-download/
Requesting access to CJARS through the FSRDC

FSRDC proposals require:
- A project overview and description of research methods
- A list of requested data and linkage procedures
- An explanation of statistical output that will be generated
- A statement of benefits of the project for the U.S. Census Bureau

Additional review by CJARS data providers depending on:
- What jurisdictions are being studied?
- What level of geographic aggregation are intended for your findings?

Contact our staff (cjars-data-users@umich.edu) with your questions or concerns
Federal Statistical Research Data Centers

Once approved, research occurs in the FSRDC network:

- Secure Census Bureau facilities housed in partner institutions
- 30 locations around the U.S.
- Access costs vary by location, affiliation, and project duration
- Researchers can bring in own data for linkage
- COVID-19 virtual access
- Only aggregate statistical material, reviewed by the U.S. Census Bureau Disclosure Review Board, can be released from FSRDCs
Researchers can link amazing data

Anonymized data can be linked at person-, address-, and employer-levels

Survey data:

Administrative records:
- Longitudinal Employer-Household Dynamics (LEHD) quarterly earnings data
- Federal program data: Medicaid and Medicare enrollment, Social Security programs, HUD assistance
- State program data: SNAP, TANF, WIC
Any views expressed herein are those of the authors and do not necessarily represent the views of the Census Bureau. All results were approved for release by the Disclosure Review Board of the Census Bureau, authorization numbers CBDRB-FY19-444, CBDRB-FY21-ERD002-017, and CBDRB-FY20-ERD002-021.
Note: Estimates based on CJARS data linked by Census-provided PIK to family relation crosswalk that was synthesized from a variety of survey and administrative data. These results were disclosed by the U.S. Census Bureau’s Disclosure Review Board with approval number CBDRB-FY20-ERD002-021.
Probability of any felony conviction by age 25

Note: Estimates based on CJARS data linked by Census-provided PIK to Census Numident (to identify Commuting Zone of birth, sex, race, and ethnicity). These results were disclosed by the U.S. Census Bureau’s Disclosure Review Board with approval number CBDRB-FY19-444.
Annual probability of employment, by prison-release cohort

Note: Estimates based on CJARS data linked by Census-provided PIK to IRS W-2 filings to measure annual employment status. Reference cohort composed of adult ACS respondents with less than a high school degree. These results were disclosed by the U.S. Census Bureau’s Disclosure Review Board with approval number CBDBRB-FY21-ERD002-017.
Research vignettes

Labor market outcomes after a criminal justice intervention
- Link CJARS individuals to Longitudinal Employer-Household Dynamics (LEHD) data to measure quarterly employment, earnings, industry

Neighborhood environment and criminal justice involvement
- Link CJARS individuals to addresses using decennial census data, measure outcomes in American Community Survey data

Criminal justice contact as an outcome for a non-CJ intervention
- Bring your person-level program participation or randomized control trial data to FSRDC, link to CJARS, measure criminal justice involvement
Integrate your data into CJARS

Continually working to expand CJARS to nationwide coverage:

- Partnerships with criminal justice agencies
- Donations of public, legally redistributable data from researchers
- Suggestions of data resources we may have overlooked

Contact cjars-staff@umich.edu to discuss how we can integrate your data holdings in future CJARS vintages
NSF-funded fellowships will jump-start access
CJARS-NSF fellowship competition

Fellowship Benefits:

1. $10,000 stipend to be used in support of your research (e.g., to defray data access costs, to support researcher time, or to hire research staff)
2. Mentorship on FSRDC proposal development from CJARS
3. Opportunity to present research findings at the annual CJARS Board of Directors meeting

Application deadline May 28, 2021. Decisions announced mid-June
How to apply

Solicitation and submission portal on project website

Application Requirements:

1. Research Proposal describing the topic of study, intended data to be linked, and research methods (2000 word limit)

2. Current C.V. or resume

3. Data Access Plan describing which FSRDC location you plan to conduct your research in, any anticipated data access barriers, and plans to address barriers (500 word limit)

Contact local FSRDC administrator about data access options prior to submission

Link: https://cjars.isr.umich.edu/2021-fellowship-competition
Proposal scoring

Four evaluation criteria:

1. Intellectual Merit of the Research Proposal
2. Compatibility of CJARS Data Holdings with Proposed Research
3. Feasibility of the Data Access Plan, and
4. PI Experience and Training

Priority groups:

1. Early stage researchers
2. Demographic groups underrepresented in Science and Engineering, and,
3. Research fields underrepresented in the FSRDC system
For more information on the material described in this webinar, please check out our website – https://cjars.isr.umich.edu – or contact the CJARS team with your questions regarding:

- Data collection/integration: cjars-staff@umich.edu
- Research projects: cjars-data-users@umich.edu
- Fellowship program: cjars-fellowships@umich.edu