Welcome!

Virtual Research Symposium Day 2
Assessing Capacity for Using Data to Build Actionable Evidence

@data_foundation | #DataLive
Opening Remarks

Dean Ritz
Chair of the Board of Directors, Data Foundation & Senior Advisor to the CEO, Open Data Policies & Practices, Workiva Inc. / @dritz63
Education Research-Practice Partnerships: Innovative Structures to Build and Use Evidence

Rachel Anderson, Director, Policy and Research Strategy, Data Quality Campaign / @rachelbarrer7
Research informs better decisions.
Research builds knowledge for the future.
Research helps teachers support every student.
Good research matters.

What happens next?
Research-Practice Partnerships: Transforming how we use data systems
WHAT IS AN RPP?
Farrell, Penuel, Coburn, Daniel, and Steup (2021)

A long-term collaboration aimed at educational improvement and equitable transformation through engagement with research. These partnerships are intentionally organized to connect diverse forms of expertise and shift power relations in the research endeavor to ensure that all partners have a say in the joint work.

KEY ELEMENTS

A long-term collaboration aimed at educational improvement or equitable transformation through engagement with research. These partnerships are intentionally organized to connect diverse forms of expertise and shift power relations in the research endeavor to ensure that all partners have a say in the joint work.
RPPs help apply data systems to pressing questions.
RPPs can help build internal data capacity.
RPPs help states and districts improve data systems.
When Researchers Have Access to Data, Students Succeed

When policymakers and researchers work together to answer questions about education, students and schools do better. That’s because research is more than an academic exercise. It creates discoveries that have a direct impact on the students of today and tomorrow. Smart state and federal investments in research and data pay off—in both innovations and student outcomes.

**RESEARCH IS ABOUT ANSWERING QUESTIONS**
Policymakers have questions about our schools. How do we keep students from dropping out? What can we do to retain the best teachers? With access to quality data and the training to use and safeguard it, researchers can help find answers and solutions.

**RESEARCH SUPPORTS INDIVIDUAL STUDENTS**
Data-driven research is behind many of the tools and strategies that help educators engage, teach, and guide students—from early warning systems to personalized learning.

**RESEARCH INFORMS BETTER DECISIONS**
Data-driven research helps states make better policy investments, like choosing an effective curriculum or developing an early childhood program that prepares students for success.

**RESEARCH BUILDS KNOWLEDGE FOR THE FUTURE**
Data-driven research helps educators and policymakers understand emerging issues and challenge conventional practices.

**HERE’S HOW RESEARCH CAN BE USED...**

**BY POLICYMAKERS**
Research helps answer policymakers’ questions and identify programs that are a good use of state funds and resources.

**BY PARENTS**
Families rely on research to help make decisions about which schools and programs are the best fit to support their child’s goals.

**BY EDUCATORS**
Research leads to effective tools, like early warning systems that help educators keep at-risk students on track.

**BY HIGHER-ED LEADERS**
Research shows postsecondary leaders how students move through their schools so they can better prepare them to graduate.
Thank you!

Rachel Anderson
Director, Policy and Research Strategy
Data Quality Campaign
randerson@dataqualitycampaign.org
Statewide Longitudinal Data Systems and Predictive Analytics: Understanding, Measuring, and Predicting K12 Outcomes

Robin Clausen, Ph.D., Stakeholder Liaison and Research Analyst, Statewide Longitudinal Data System, Montana Office of Public Instruction

Angie Henneberger, Ph.D., Research Assistant Professor, University of Maryland School of Social Work / @mdsocialwork

Nancy Sharkey, Ed.D., Senior Program Officer, Statewide Longitudinal Data Systems Grant Program, National Center for Education Statistics / @EdNCES

Chris Stoddard, Ph.D., Professor, Montana State University / @montanastate

Mathew Uretsky, Ph.D., Professor, Portland State University / @pdxssw
Statewide Longitudinal Data Systems and Predictive Analytics: Understanding, Measuring, and Predicting K12 Outcomes

June 23, 2022

Nancy Sharkey
SLDS Program Officer

Robin Clausen
Montana Office of Public Instruction

Chris Stoddard
Montana State University

Mathew Uretsky
Maryland Longitudinal Data System Center

Angela Henneberger
Maryland Longitudinal Data System Center
Agenda

• SLDS Grant Program Overview
• State Perspective: Montana
• State Perspective: Maryland
SLDS Program Background
Legislative Background

- The SLDS Grant Program was authorized in 2002 by the Education Sciences Reform Act and the Educational Technical Assistance Act.
- The grants are cooperative agreements, which have more active federal government involvement than typical grants.
- Grants are administered by the Institute of Education Sciences (IES) of the U.S. Department of Education.
Grant Awards

To date, 49 states plus American Samoa, the Commonwealth of the Northern Mariana Islands, the District of Columbia, Guam, Puerto Rico, and the U.S. Virgin Islands have received grants totaling $826 million in 7 rounds of grants.
Program Evolution

2006 & 2007

K12

Number of Grants: 14, 13
Avg. Award: $3.7M & 4.8M

2009

K12 + ONE of the following:
• PreK
• Postsec.
• Workforce
OR
• Student-teacher link

27
$5.6M

2009 ARRA

K12 + ALL of the following:
• PreK
• Postsec.
• Workforce
AND
• Student-teacher link

20
$12.5M

2012

ONE of the following:
• PreK
• Postsec.
• Workforce
OR
• Postsec/Workforce

24
$4.1M

2015

ONE or TWO of the following:
• Financial equity and return on investment
• Educator talent management
• Early learning
• College and career
• Evaluation and research
• Instructional support

16
$6.5M

2019

ONE of the following:
• Infrastructure
• Education choice
• Equity
OPTIONAL
• School-level poverty measure

28
$3.3M
FY19 Grants: Key Information

- Length of grants: 4 years
- Start and end dates: March 1, 2020 – February 28, 2024
- Three Priority Areas: Infrastructure, Education Choice, Equity

<table>
<thead>
<tr>
<th>Priority Area</th>
<th>Number of Grants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>19</td>
</tr>
<tr>
<td>Education Choice</td>
<td>1</td>
</tr>
<tr>
<td>Equity</td>
<td>8</td>
</tr>
</tbody>
</table>
SLDS Technical Assistance: State Support Team
The EDTAP Contract
SLDS State Support Team (SST)

- Free technical assistance is available to all states and territories, regardless of whether or not they have an active SLDS grant.
- Each state and territory has an assigned SST Point of Contact.

SST State Point of Contact
One Pager:
What Is the State Support Team?

The State Support Team is a group of data systems experts whose primary objective is directly supporting states to develop, manage, use, and sustain SLDSs.
What Is the State Support Team?

Some common areas of SST assistance include:

- data governance
- stakeholder engagement
- project management
- sustainability
- data use
- transparency
- collaboration and data linking
- vendor management
- system design
- strategic planning
- research agendas
What Is the State Support Team?

The SST provides support in sectors critical to SLDS success:

- early childhood initiatives
  - K12 systems
    - postsecondary systems
      - workforce data
      - social services data
Types of SST Assistance

INDIVIDUAL ASSISTANCE

• Check-in calls and on-site visits
• Review project plans and other documents
• Provide resources, training, and examples
• Connect states with their peers
• Help determine state needs
Types of SST Assistance

INDIVIDUAL ASSISTANCE
• Check-in calls and on-site visits
• Review project plans and other documents
• Provide resources, training, and examples
• Connect states with their peers
• Help determine state needs

MULTI-STATE ASSISTANCE
• Best Practices Conference
• Resource sharing repository
• Publications
• Topical working groups
• Webinars
• Communities of practice
• Personnel Exchange Network
• Listserv
SLDS Best Practices Resources

The SLDS team produces a variety of products to capture best practices from the field and meet the evolving needs of the community.

- Issue Briefs
- Guides
- State/Territory Spotlights
- Webinar Summaries

Search or browse resources at https://slds.ed.gov/#p=19.
Communities of Practice

- SLDS Communities of Practice are based on the [SLDS Framework](#).
- The framework covers seven essential components of an effective SLDS.
- The SST organizes all its technical assistance around these components.
Webinars

Webinars provide venues for states to:

• demonstrate products they have developed
• share best practices
• discuss issues of interest
• share valuable knowledge in a timely fashion

All webinars are recorded, and many have webinar summaries available on Communities360.

Webinars are open to staff from all state agencies.
SST Collaboration: Integration Tier Partners

- Common Education Data Standards (CEDS)
- Privacy Technical Assistance Center (PTAC)
- Center for the Integration of IDEA Data (CIID)
- EDFacts Partner Support Center and EDFacts Data Management Team
The Common Education Data Standards (CEDS)

The CEDS project is a national collaborative effort that has developed voluntary, common data standards to streamline the exchange, comparison, and understanding of data within and across P-20W institutions and sectors.

- CEDS Data Warehouse
- CEDS Elements
- CEDS Integrated Data Store

http://ceds.ed.gov
The Common Education Data Standards (CEDS) Tools

• Align
• Connect
• Open Source Community

http://ceds.ed.gov
Help with The Common Education Data Standards (CEDS)

Attend a CEDS Open Source Community Meeting!
• Occur the first Monday of every month
• 2:00-3:00 PM ET

The CEDS Open Source Community is where all changes to CEDS begin. New elements, element updates, and improvements to the CEDS Integrated Data Store and CEDS Data Warehouse are processed through the CEDS Open Source Community.

http://ceds.ed.gov
Collaboration with Partner Support Center and Data Management: Value to Staff

- SST will know the current status, plans, and milestones for modernization planning and implementation so they can reference it and related resources in their SLDS technical assistance to states.
- EDFacts PSC staff will have a greater understanding of the State Support Team scope and delivery of technical assistance, which will enable state referrals to SST when SEAs have broader organizational and data quality challenges that impede EDFacts reporting.
- EDFacts PSC and DM will develop greater understanding of the current state of SEAs’ human, organizational, and technical capacity, how federal reporting is currently conducted, and the areas of potential challenge universally and individually as modernization progresses.
Collaboration with Partner Support Center and Data Management: Value to States and Territories

- States and territories will receive more holistic technical assistance which addresses their SLDS and federal reporting work in relation to one another.
- They will be better prepared to proactively incorporate the requirements of modernization into their internal infrastructure modernization efforts, including planning for the needed changes in data governance and specific processes.
Contact Information and Additional Information

- Nancy Sharkey nancy.sharkey@ed.gov
- About the SST: https://slds.ed.gov/#program/about-the-sst
- About the SLDS Grant Program: https://nces.ed.gov/programs/slds/
State Perspective: Montana

Robin Clausen, Montana Office of Public Instruction
Chris Stoddard, Montana State University
EVALUATING Montana’s Early Warning System (EWS)

Dr. Robin Clausen
Office of Public Instruction
Dr. Chris Stoddard
Montana State University

Putting Montana Students First
Online EWS Tools

• School level report - Summarizes data and creates visualizations for school level dropout risk, and specific trends including grades, attendance, behavior, and mobility.

• Student summary report - Generates a spreadsheet containing all student data for the school, including risk rankings, percentage risk, change in risk, and odds ratios for specific risk factors.

• Student detail report - Provides data and visualizations for a single student within that school, including their current dropout risk, change in risk over time, information on missing data, and predominant risk factors where interventions may be warranted.

• Dropout Probability- In grade 9-12 an at-risk student is identified as having a > 15% probability to drop out. Extreme at-risk student have a > 40% probability.
### Student Level Report

**Student Name:** Jess Thompson - UDJEHEGDB

#### Dropout Probability

**Dropout Probability:** 60.2%

#### Dropout Risk Factors

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older Student</td>
<td>Y</td>
</tr>
<tr>
<td>Off Track</td>
<td>N</td>
</tr>
<tr>
<td>Previous Dropout</td>
<td>N</td>
</tr>
<tr>
<td>Attendance Risk Factor</td>
<td>1.34</td>
</tr>
<tr>
<td>Grades Risk Factor</td>
<td>1.00</td>
</tr>
<tr>
<td>Behavior Risk Factor</td>
<td>1.34</td>
</tr>
<tr>
<td>Mobility Risk Factor</td>
<td>1.00</td>
</tr>
</tbody>
</table>

#### Dropout Probability Summary

<table>
<thead>
<tr>
<th>Date</th>
<th>Early Warning System</th>
<th>Dropout Probability</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 Oct 2016</td>
<td>60.2%</td>
<td>62.1%</td>
<td></td>
</tr>
<tr>
<td>21 Oct 2016</td>
<td>61.0%</td>
<td>62.1%</td>
<td></td>
</tr>
<tr>
<td>21 Apr 2016</td>
<td>74.3%</td>
<td>73.5%</td>
<td></td>
</tr>
<tr>
<td>19 Apr 2016</td>
<td>67.5%</td>
<td>73.5%</td>
<td></td>
</tr>
<tr>
<td>17 Mar 2016</td>
<td>70.4%</td>
<td>76.3%</td>
<td></td>
</tr>
<tr>
<td>05 Mar 2016</td>
<td>73.3%</td>
<td>76.3%</td>
<td></td>
</tr>
<tr>
<td>04 Mar 2016</td>
<td>76.3%</td>
<td>81.5%</td>
<td></td>
</tr>
<tr>
<td>17 Nov 2015</td>
<td>76.4%</td>
<td>81.5%</td>
<td></td>
</tr>
<tr>
<td>14 Oct 2015</td>
<td>60.2%</td>
<td>81.5%</td>
<td></td>
</tr>
<tr>
<td>23 Sep 2015</td>
<td>61.5%</td>
<td>81.5%</td>
<td></td>
</tr>
<tr>
<td>22 Sep 2015</td>
<td>61.5%</td>
<td>81.5%</td>
<td></td>
</tr>
</tbody>
</table>

#### Dropout Probability Graph

- **Graph Title:** Dropout Probability
- **Graph Data Points:**
  - 23 Jun 2015: 55.0%
  - 30 Jun 2015: 60.0%
  - 29 Jul 2015: 65.0%
  - 22 Sep 2015: 75.0%
  - 21 Apr 2016: 85.0%

#### Grades Risk Factor Graph

- **Graph Title:** Grades Risk Factor
- **Graph Data Points:**
  - 23 Jun 2015: 0.9
  - 30 Jun 2015: 1.0
  - 29 Jul 2015: 1.0
  - 22 Sep 2015: 1.3
  - 21 Apr 2016: 1.2

#### Attendance Risk Factor Graph

- **Graph Title:** Attendance Risk Factor
- **Graph Data Points:**
  - 23 Jun 2015: 0.0
  - 30 Jun 2015: 0.0
  - 29 Jul 2015: 0.0
  - 22 Sep 2015: 0.0
  - 21 Apr 2016: 0.0
Research question 1: Does EWS Accurately predict graduation?

- Early Warning System model was based on pilot school data
- Currently uses attendance, grade retention, moves across schools, behavior incidents (suspensions, expulsions) to predict dropout probability.
- 15% or greater flagged as “At Risk”, 40% or greater “Extreme at Risk”

<table>
<thead>
<tr>
<th></th>
<th>EWS predicted graduation</th>
<th>EWS predicted dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual high school graduate</td>
<td></td>
<td>“False negatives”</td>
</tr>
<tr>
<td>Actual high school dropout</td>
<td>“False positives”</td>
<td></td>
</tr>
</tbody>
</table>

- “False Negatives” are a success! School interventions based on EWS may have prevented Dropout
- Analysis will focus on “False Positives”
Research Question 2: What is the degree of implementation of ews model in participating schools?

○ We know a great deal about the implementation of the pilot schools (18)
○ Know less about how the other 122 schools are using the system.
○ Surveys and interview school leaders in Montana in schools that participate in EWS. Surveys were distributed in Spring 2022 by the Montana Office of Public Instruction.
○ Create an implementation index (on a scale from 1-4)
○ We will refine this index with the results of 45 interviews conducted with school officials in Year Two of the research study.
Research Question 3: Does the EWS improve student outcomes?

- Examine attendance, high school graduation, and college attendance
- Use staggered rollout of EWS to trace out effects
  - Compare students in schools before and after EWS was adopted
  - Compare students in EWS schools to students in schools not using EWS
  - Compare students who were “exposed” to EWS longer than others. Students graduating in 2012 only “exposed” for 1 year, by 2020 exposed since elem/middle school

Does the EWS improve student outcomes for specific subgroups of students?
- Gender, race/ethnicity, school size, school locale, intensity of EWS use
- Detailed analysis focusing on students identified as “at risk”
Getting to know the data
Our task: Focus on Using SLDS Data and On Implementation

• MT OPI has delivered the first round of data to MSU. This round included records on all students since 2008 and related datapoints about EWS students since 2011.
• MSU has completed tasks for Year One of the NCER research study. While results are preliminary, they highlight challenges and opportunities with the data management and analysis. Year Two of the study for MSU will focus on research question 1 and 3.
Schools’ use of EWS varied considerably
How well did EWS predict final dropout rates?

<table>
<thead>
<tr>
<th>Of students observed long enough to see final graduation outcome</th>
<th>Ever Graduated</th>
<th>Dropped out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students ever scored at extreme risk of dropping out (N=4,115)</td>
<td>42.1%</td>
<td>57.9%</td>
</tr>
<tr>
<td>Students ever scored at risk of dropping out but never at extreme risk (N=5,465)</td>
<td>62.6%</td>
<td>37.4%</td>
</tr>
<tr>
<td>Students never flagged as at risk (N=14,110)</td>
<td>92.3%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>
What do these preliminary results indicate?

○ EWS model is strongly correlated with actual graduation experiences.
  ○ Very few students never flagged by the system as at risk ever drop out.
  ○ Predicted probability of dropout is strongly related to actual dropout.

○ Students without an EWS score have higher dropout rates than students in the EWS system.

○ Important caveat: non-participating schools might be different than participants. Maybe non-participants have fewer counselors, less resources, etc. and therefore would have had higher dropout rates regardless of EWS use.

○ Future research will compare how dropout rates changed for EWS adopters and non-adopters over time to distinguish these possibilities.
State Perspective: Maryland

Mathew Uretsky, Portland State University, Maryland Longitudinal Data System Center
Angela Henneberger, University of Maryland School of Social Work, Maryland Longitudinal Data System Center
Persisting Students as a Unique Group of High School Non-Graduates in Maryland

Mathew C. Uretsky¹
Angela K. Henneberger²

¹Portland State University
²University of Maryland School of Social Work

2022 Virtual Research Symposium: Assessing Capacity for Using Data to Build Actionable Evidence

June 23, 2022
Overview

○ Introduction to the MLDS
○ Introduction and background on persisters
○ Identifying persisters in Maryland
○ Rates of persisters in Maryland
○ Outcomes for persisters in Maryland
○ Summary
○ Implications
The MLDS Center

- **Independent state agency** that develops and maintains a data system containing student and workforce data from all levels of public education and the State’s workforce.

- **Generate timely and accurate information** about student performance that can be used to improve the State’s education system and guide decision makers at all levels.
Introduction and Background on Persisters

- Most studies focus on dropouts
  - Early warning indicators or
  - Typologies (e.g. pushouts, pullouts)
- Who is missing from the conversation?
  - What about students who don’t dropout, but do not graduate on time?
Useful Definitions

○ **On-time graduate** – Students who graduate with a regular diploma four years after entering the ninth grade as first-time freshmen

○ **Dropout** – Students who formally withdraw from school

○ **Persister** – Students who do not earn a regular diploma but are still enrolled on their expected graduation date
Identifying Persisters in Maryland
Establishing the Cohort

Year 0
• Establish 8th Grade Enrollment

Year 1
• Establish 9th Grade Enrollment

Year 4
• Establish Exit Status
Figure 1.
Systems Diagram for Categorizing High School Graduation Outcomes
Persisting Students in Maryland

Rates and Outcomes
Sankey Diagram Years 4-6

<table>
<thead>
<tr>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Graduates</td>
<td>4,962</td>
<td>College</td>
</tr>
<tr>
<td>Year 4 Persisters</td>
<td>3,991</td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 4 Dropouts</td>
<td>971</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| | Total (N=4,962) | Year 4 Graduation Status |
| | | Persister (n=3,997) | Dropout (n=965) |
| | % | % | % |
| Post Y4 Enrollment | | | |
| Total | 81 | 97 | 14 |
| Summer | 80 | 38 | <1 |
| Year 5 | 20 | 81* | 14 |

| | No High School Credential | No Participation |
| | 58% | 64% |
| | 2,853 | | |
| | | | |

66
Sankey Diagram Years 4-6

Year 4 Per:
- Non-Graduates: 3,991 (80%)
- Year 4 Dropouts: 971 (20%)

Year 5 Graduation Status:
- Graduate: 38%
- GED: 4%
- No HS Credential: 58%

Year 4 Graduation Status:
- Total: (N=4,962)
  - Persisters: 47%
  - Dropouts: 4%

Year 6:
- College: 10%

Year 6 Per:
- 472
- 1,321
- 3,169
- No Participation: 64%
## Sankey Diagram Years 4-6

### Year 4
- Non-Graduates: 4,962
- Year 4 Persisters: 3,991 (80%)
- Year 4 Dropouts: 971 (20%)

### Year 5
- SUNY: 3,997 (81%)

### Year 5 Graduation Status

<table>
<thead>
<tr>
<th>Graduation Status</th>
<th>Year 5 Graduates (n=1,901)</th>
<th>GED (n=208)</th>
<th>No HS credential (n=2,853)</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

| Total             | 38%                         | 4%          | 58%                       |
| College           | 18%                         | 21%         | 3%                        |
| Worked 3Q+        | 38%                         | 44%         | 23%                       |
| No Participation  | 47%                         | 45%         | 75%                       |

### Year 6
- College: 472 (18%)
- No Participation: 3,169 (64%)
- No High School Credential: 2,853 (58%)
Postsecondary and Workforce Outcomes by High School Completion Status through Year 11 (N= 54,023)
Median Annual Wages by High School Completion Type Calendar Years 6-11
Summary and Implications
Summary

○ Persisting students
  ○ Consistently outnumber dropouts (4:1 in Year 4)
  ○ Higher reenrollment (>7:1)
  ○ Higher Year 5 graduation rate (>12:1)

○ Nearly all persisters reenroll, but more than half have no HS credential by the end of Year 5

○ GED earners and Late grads
  ○ Similar post-secondary and workforce participation
  ○ Late grads out-earn GED earners by Year 11
**Implications**

- Considering persisting *alongside* dropout provides a more informative analysis of high school graduation.
- Better identification and continued study of persisters:
  - Can help to increase high school graduation rates for persisters and dropouts.
  - And promote a cascade of positive life outcomes.
- Encouraging late graduating or GED earning may present a viable alternative.
- Population-level linked longitudinal administrative data allows for investigation of relatively rare populations.
For More Information

○ MLDS Center website
  https://mldscenter.maryland.gov/

○ Reports and publications available upon request – Email:
  muretsky@pdx.edu
Questions and Contact

Dr. Mathew Uretsky
Portland State University
MLDS Center
muretsky@pdx.edu

Dr. Angela Henneberger
University of Maryland School of Social Work
MLDS Center Director of Research
angela.henneberger@maryland.gov
Contact Information and Additional Information

SLDS Program
• Nancy Sharkey: nancy.sharkey@ed.gov

Montana
• Robin Clausen: robin.clausen@mt.gov
• Chris Stoddard: chrisstodd@gmail.com

Maryland
• Mathew Uretsky: muretsky@pdx.edu
• Angela Henneberger: angela.henneberger@maryland.gov

About the SLDS Grant Program: https://nces.ed.gov/programs/slds/
About the SST: https://slds.ed.gov/#program/about-the-sst
Questions and Discussion
Thank you!
5-minute Break

Event will resume at 11:30 a.m. ET
Using Linked Administrative Data to Connect Families to Pandemic Stimulus Payments

Aparna Ramesh, Senior Research Manager, California Policy Lab, UC Berkeley / @arparnar
Connecting families to benefits using administrative data: a case study in hashed linkages

Aparna Ramesh, Senior Research Manager
California Policy Lab at UC Berkeley

aparna.ramesh@berkeley.edu

June 2022
The California Policy Lab

• Research institute at Univ of California, that works almost exclusively with admin data
• Works with 40+ state and local gov’t partners
• Data on…
  • Wages, taxes, and UI
  • Human services and homeless-system enrollments
  • Health utilization and vital records
  • Arrests, charges, and dispositions
  • K12 and higher ed enrollments and outcomes
• Safety net research primarily focused on:
  • How to equitably increase access?
  • How to improve service delivery?
  • What is the impact of the safety net on health, employment, education, criminal justice?
The policy problem

• Anti-poverty tax credits: cash delivered through tax filing with potential to lift millions of out poverty

• federal Earned Income Tax Credit, federal stimulus payments, advanced Child Tax Credit (plus state versions): thousands of dollars in aid

• Many who qualify actually fall below the filing threshold and may not regularly file taxes: and miss out on thousands in benefits

• Three vexing policy questions:
  • How many low-income Californians qualify for but are not receiving this cash assistance (aka who are the non-filers)?
  • How can we close the take-up gap?
  • If they receive these credits, how many Californians are lifted out of poverty?
The data challenge: no one entity knows who the non-filers are

Tax agencies (e.g. IRS, California Franchise Tax Board):
- Know who files and receives credits.
- Has individual non-filer info
- Do not know who is eligible but non-filing.

State and local human service agencies (e.g. CA Dept of Social Services):
- Serve many families below the non-filing threshold.
- Do not know which of those families have not filed.

Take-up:

Number of eligible non-filers

All eligible individuals
The privacy challenge

Agency 1

“I don’t want to share my PII”

Agency 2

“I don’t want to share my PII”

Rationales

• Legal
• Privacy
• Security
• Trust
Privacy preserving record linkage (using third-party)
Our use case

Goal

- To identify enrollees of SNAP that were eligible for — but not claiming — tax benefits like EITC, CTC, and stimulus $
The solution: a hashed linkage

- One-directional encryption
- Mathematically impossible to reverse
- Hash SSN, first name, last name, and date of birth

```
PII (SSN)       Hash algorithm     Unique hash value
123-45-6789     SHA 256            01a54629efb952287e554eb23ef69c52
                 SHA 256            097a75aecc0e3a93ca0855ab6d7a31a0
123-45-6780     SHA 256            87dbe22abba0e202c06b4671f06d08a2
                 SHA 256            dbcb0a7c095019f65ef11a06a00ed541
```

Hash is completely different if even ONE digit in SSN changes
Serving as a third party

- CPL serves as a trusted third party: has a DUA signed with both FTB and CDSS—CDSS/FTB do not need to sign anything with each other
- Hashing is a one-way transformation of identifiers in the same way—so the receiving agency doesn't know but can link the data

Tax filing data  
(FTB: 2018 + 2019)  
Hashed

CPL merges data

Generates list of CDSS clients who have not filed a return

Send back to CDSS, which merges with reidentified data

Safety net data  
(CDSS: 2018-2020)  
Hashed
What we found

Most low-income families actually do file a return and received benefits automatically

75% received the federal stimulus automatically

74% of children received child tax credit automatically
What we found

Of the adults who did not receive the federal stimulus:

- Most were single adults without dependents (60%)
- Most had no observed wage earnings (67%)

Of the children who did not receive the credit:

- Most live with a single adult
- A quarter lived in households with no adult on the case
So what?
“Finding a non-filer is like finding a needle in a haystack”
CDSS used the non-filer list to conduct outreach

- CDSS conducted phone and email outreach to roughly ~400,000 non-filers
- Directed them to the non-filer portal (GetCTC.org) with hotline assistance
- We built a randomized control trial, currently using a hashed linkage to evaluate who eventually filed a return
- What worked? What didn’t? What are higher-touch interventions to help individuals file tax returns?
Want to administer hashed linkage?

Visit: bit.ly/HashedLinkage

Contains

- **Guide for departmental leadership:** to decide on the best legal pathways to linkage, assemble and manage a team of legal and data staff to execute a linkage, and identify and work with a third-party partner as needed

- **Step by step guidance for technical staff:** to execute a privacy-preserving linkage using cryptographic hashing

- **Hashing code on GitHub** (linkage code coming soon!)  
  https://github.com/californiapoliciylab/hashed-linkage
Beyond this policy problem

• Social services <> colleges/universities <> financial aid
• Social services <> incarceration
• Social services <> credit bureau
• Social services <> public utilities <> credit bureau
• Social services <> birth/death records <> hospitalizations/ED
• Credit bureau <> financial aid
• Credit bureau <> probation
Takeaways:

#1: Breaking down data silos can help break down policy silos

#2: Hashed linkages can preserve privacy and break silos

#3: Linked data can help identify the problem and evaluate the solution: building a virtuous cycle of data use
Questions?
aparna.ramesh@berkeley.edu
bit.ly/HashedLinkage
Using a Framework for Evidence Capacity to Strengthen Federal Program Offices

Heather Gordon, Managing Consultant, Mathematica / @MathematicaNow
Using a Framework for Evidence Capacity to Strengthen Federal Program Offices

Heather Gordon, Mathematica
How do we define evidence capacity?

Evidence capacity encompasses the knowledge, skills, behavior, and resources that support an organization’s ability to build and use evidence to make decisions.
**Fast facts about the framework**

/ **Who it’s for**  
OPRE, but is applicable for broader use across the federal government and human service organizations

/ **Why it’s needed**  
To establish a practical definition of “evidence capacity” and assess an organization’s evidence capacity

/ **Where to find it**  
Will be released publicly on OPRE’s website later this year
Framework structure

**Dimension:** a grouping of evidence capacity components based on a common theme

**Component:** an input, output, and/or activity that is involved in evidence capacity building
Five dimensions of evidence capacity

1. **Evidence culture**
   - Organizational commitment to evidence
   - Equity and inclusion
   - Learning mindset

2. **Leadership**
   - Evidence-informed decision-making
   - Budgeting
   - Team composition
   - Team support
Five dimensions of evidence capacity cont.

3. **Evidence infrastructure**
   - Tools
   - Data infrastructure and technology
   - Performance monitoring and improvement
   - Program evaluation

4. **Human capital**
   - Evidence-building and analytic skills
   - Communication skills
   - Professional development

5. **Engagement**
   - Dissemination
   - Internal engagement
   - External engagement
5 key dimensions of evidence capacity
Office of Human Services Emergency Preparedness and Response (OHSEPR) partnership

Goal: Support OHSEPR’s evidence capacity by providing foundational information on disaster displacement, an issue that is not well addressed in existing resources

Relevant evidence capacity components:
- Organizational commitment to evidence
- Learning mindset
- Evidence-informed decision-making
- External engagement
Goal: Improve the performance monitoring capacity within the SOT program by using existing data to answer specific research questions and identify ways to improve existing measures

Build the foundation for future evaluation to determine what works in terms of SOT program characteristics and client outcomes

Relevant evidence capacity components:
- Organizational commitment to evidence
- Learning mindset
- Performance monitoring and improvement
- Data infrastructure and technology
- External engagement
For more information, please contact:

/ **ACF/OPRE/Division of Data and Improvement:** Nicole Deterding (Nicole.Deterding@acf.hhs.gov)

/ **Mathematica:** Heather Zaveri (hzaveri@mathematica-mpr.com) and Heather Gordon (hgordon@mathematica-mpr.com)
Questions?
30-minute Break
Event will resume at 12:30 p.m. ET
Welcome back!

12:35 p.m. – Critical Factors for Building Successful Data Science Teams

12:50 p.m. – Advocating for and Applying COVID-19 Equity Data: The Black Equity Coalition’s (Pittsburgh, PA) Efforts to Improve Public-Sector Health Agencies' Practices

1:50 p.m. – 5-minute break

1:55 p.m. – A dynamic, inclusive approach to learning agenda development for the Centers for Disease Control and Prevention’s (CDC’s) Center for State, Tribal, Local, and Territorial Support (CSTLTS): Reflections on the participant engagement process

2:10 p.m. – Best Practices for Monitoring and Evaluating the ARP, IIJA and Other Programs: Report of the Department of Commerce Data Governance Working Group

3:10 p.m. – Closing remarks

@data_foundation | #DataLive
Critical Factors for Building Successful Data Science Teams

Robin Wagner, Ph.D., Senior Advisor, Epidemiology Branch, Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, National Institutes of Health / @nih_nhlbi
Critical Factors for Building Successful Data Science Teams

Robin M. Wagner, PhD, MS
Chair, HHS Data Council Data-Oriented Workforce Subcommittee (DOWS) and
Senior Advisor, Division of Cardiovascular Sciences
National Heart, Lung, and Blood Institute
National Institutes of Health

Presented to Data Foundation 2022 Virtual Research Symposium:
Assessing Capacity for Using Data to Build Actionable Evidence
June 23, 2022
Overview

• Background
• Study Purpose
• Overall Study Design
• Literature Review Key Findings
• Gaps in Literature Review and Solutions
• Final Summary Report Recommendations
• Discussion
Data-Oriented Workforce Subcommittee (DOWS)

- Subcommittee of HHS Data Council
- Established to implement workforce priority of HHS Data Strategy

**CHARGE**

Develop high-level work plan to enhance the data science capacity of HHS’ workforce by identifying training opportunities for existing staff, recruitment strategies and tools to hire new staff, and retention and succession planning strategies to sustain the data science workforce

- Focus on low-hanging fruit that benefits all HHS agencies, but consider longer term strategies
- Include strategies that build data science capacity at organization, team and individual levels
- Align work with Foundations of Evidence Based Policymaking Act of 2018, Federal Data Strategy, HHS Data Strategy, and HHS’ agencies’ data strategies
DOWS Progress to Date

31 members from 11 HHS agencies

44 meetings since June 2019

35+ presentations/demos from HHS, other agencies

Case study approach to synthesize findings across agencies

Initial assessment of fellowships/internships that can support data science

Critical Factors for Building Successful Data Science Teams

Authorities and Mechanisms for Hiring and Retaining Data Scientists at HHS
Critical Factors for Building Successful Data Science Teams
Study Purpose

Create evidence

• To identify critical factors for building successful data science teams and optimizing their use and value within their organizations

Disseminate findings

• To provide evidence-based recommendations to the HHS Data Council for wider dissemination to HHS OPDIVs and STAFFDIVs and other federal partners
Overall Study Design

Initial 9 policy questions

Three pre-literature review focus groups:
- Two focus groups with members in the data scientist community
- One focus group with members of the Federal Chief Data Officer Council

Three key informant interviews with senior, experienced data science leaders

12 revised policy questions

Literature review

Final summary report with recommendations on building successful data science teams in the twenty-first century

One focus group with members of the data science community to provide feedback on the literature review

Literature review summary
Literature Review Key Findings

**Topic 1: Skills and competencies of data science teams**
- Data science skillsets fall into technical, analytical, and business domains, while data ethics should overarchingly guide data science work
- Organizations should distribute skills among data science team members, since unicorns are rare, command high salaries, and do not mitigate risk

**Topic 2: Management**
- Data science managers must be data literate, need to understand how data science can help the organization achieve its goals, and should foster an environment for data-driven experimentation

**Topic 3: Recruitment and retention**
- Data scientists are looking for interesting and challenging work and a place to grow, so an environment that offers them career development opportunities and clear career pathways, fosters creativity and curiosity, and flexibility is key

**Topic 4: Organization of data science teams**
- A hub-and-spoke model, in which a centralized hub exists that connects to other departments, or spokes, is the most recommended organizational model for data science teams

**Topic 5: How organizations value and use data science teams**
- Organizations need strong communication, clear understanding and expectations of their data science team(s), a shared vision for data science projects, sufficient IT resources, and a culture of data literacy to ensure data science teams are used and valued by their organization
Gaps in Literature Review and Solutions Identified by Focus Group

• Evaluate data scientist skills
  – Use take-home data science exercises and diverse business areas hiring panel for candidates

• Structure data science teams in organization
  – Vary team structure based on organization’s needs

• Address federal government-specific challenges
  – Recruitment: Long hiring processes, unclear job announcements and expected career paths, lower salaries compared to private sector
  – Retention: Poorly defined data science projects, inadequate data science resources/tools, inadequate organizational data literacy and democratization (requires leadership support)

• Enhance data ethics including fairness/transparency/checks into daily work

• Enable data scientists to serve as agency advocates and change agents

• Provide more concrete examples to facilitate agency transformation

• Expand research on Chief Data Officer’s role in enhancing federal data science workforce
Final Summary Report Recommendations – I

Topic 1: Skills and competencies of data science teams
- Distribute skills across team members
- Regularly measure and upgrade skills within the data science team
- Ensure team members have diverse backgrounds and an understanding of how data ethics fits into their work
- Develop skills from the technical, analytics, and business domains within the data science team

Topic 2: Management
- Ensure data science team managers are data literate with strong business skills and awareness of the goals of the broader organization
- Enable data science managers to serve as advocates for the data science team to upper management

Topic 3: Recruitment and retention
- Provide challenging and mission-oriented work to attract and retain talented data scientists
- Provide clear career pathways and professional development opportunities to retain data scientists
- Work with hiring managers to be intentional in the language and expectations listed in job announcements
- Use new hiring processes such as Subject Matter Expert Qualification Assessments and develop partnerships with universities to recruit data scientists
Final Summary Report Recommendations – II

**Topic 4: Organization of data science teams**
- Use a hybrid of centralized and decentralized approaches, also known as a hub-and-spoke approach, to organize data science teams within the organization
- Ensure the hub is led by executive leadership to maintain visibility within the organization
- Create a Center of Excellence in the hub to provide data scientists an identity within the organization
- Create a structure that provides opportunities for experimentation and exploration so that the data scientists can thrive

**Topic 5: How organizations value and use data science teams**
- Secure buy-in from executive leadership who agree to promote a culture that welcomes new data scientists, ideas, and technologies
- Increase data literacy throughout the organization to facilitate strong communication between data scientists and end users
- Implement policy changes to democratize data and integrate data across multiple agencies
- Enlist the support of executives and collaborate with organizational groups and end users to empower data scientists to serve as advocates for change within their organization
Discussion
Robin M. Wagner, PhD, MS
Senior Advisor
Epidemiology Branch, Division of Cardiovascular Diseases
National Heart, Lung, and Blood Institute
Phone: (301)-480-7019
Email: Robin.Wagner@nih.gov
Advocating for and Applying COVID-19 Equity Data: The Black Equity Coalition's (Pittsburgh, PA) Efforts to Improve Public-Sector Health Agencies' Practices

Jason Beery, Ph.D., Director of Applied Research, UrbanKind Institute & Member, Black Equity Coalition Data Working Group / @BeUrbanKind

Ashley Hill, DrPH, Assistant Professor, Department of Epidemiology, School of Public Health, University of Pittsburgh & Member, Black Equity Coalition Community Health Working Group / @DrAshleyHill

Ruth Howze, Community Coordinator, Black Equity Coalition / @blackequitypgh

Stacey Pharrams, Community Researcher, Healthy Start Initiative & Member, Black Equity Coalition Data Working Group
Advocating for and Applying Equity to COVID-19 Data:

The Black Equity Coalition's (Pittsburgh, PA) Efforts to Improve Public-Sector Health Agencies' Practices

Stacey Pharrams, Community Researcher, Healthy Start Incorporated
Ruth Howze, Community Coordinator, Black Equity Coalition
Ashley V. Hill, DrPH, Epidemiology, University of Pittsburgh
Jason Beery, PhD, Director of Applied Research, UrbanKind Institute

Data Foundation 2022 Research Symposium
June 23, 2022
The Black Equity Coalition

The Black Equity Coalition (BEC) is a Black-led multi-sectoral collaboration which aims to address the disparate impact of the COVID-19 pandemic on the Pittsburgh region’s Black community.

Members are:

- physicians
- epidemiologists
- public health and health care practitioners
- social scientists
- business leaders and business support providers
- community funders
- government officials
- civic data intermediaries
BEC Data Committee
Administrative context

- Allegheny County Health Department is the legislatively-designated public health authority.

- Across Pennsylvania, five other counties and four municipalities have their own public health authorities.

- The PA Department of Health manages the remainder of the state (~59% of PA population).

- Exception: Philadelphia’s health department is entirely independent from the PA Department of Health.

https://www.health.pa.gov/About/Pages/County-Municipal%20Health%20Depts.aspx
In April, 2020, a group of people voluntarily came together to look at data to make sense of the pandemic

- Karen Abrams, Program Officer at The Heinz Endowments pulled together epidemiologists, health professionals, social scientists, community leaders, and data professionals.
- The group broadly explored how the pandemic was impacting the Black community, and began meeting twice per week.
- In time, the group became the Data Committee of the Black Equity Coalition.

You need to come to this meeting
Goal: Improve access to testing

- Early-on, it was clear that testing resources weren’t accessible to people in many Black communities

- BEC shared maps showing this disparity with the Pittsburgh Black Elected Officials Coalition
Outcome: Allegheny County Health Department began providing testing through Federally-Qualified Health Centers
Goal: Improve quality of testing and case data

- **Testing and case records were missing race and ethnicity.**
- **Asked county to:**
  - link case data to their data warehouse to fill-in missing values.
  - emphasize collection of data by race in contact tracing, and to hire contact tracers through FQHCs to improve response by Black people.
- **Encouraged PA Department of Health to provide training and emphasize enforcement of data collection practices.**

https://www.health.pa.gov/topics/disease/coronavirus/pages/cases.asp
PA DoH emphasizes data quality improvements through orders and training
Outcome: A growing share of case records from Allegheny County and Pennsylvania capture race

March 2021: 88%
October 2021: 94%

March 2021: 63%
October 2021: 73%
Goal: Improve accessibility of data - We started with a daily email of data scraped from public websites...

Here are the state and county updates for 4/26:

**FA State COVID-19** data from 2020-04-26:
- Positive Cases: 31.2% Black (4027 Black / 12915 race known, 28250 race unknown, 68.6% of 41165 total)
- Deaths: 22.0% Black (166 Black / 763 race known, 787 race unknown, 50.0% of 1550 total)
- Positive Cases: 64.2% White (8292 White / 12915 race known, 28250 race unknown, 68.6% of 41165 total)
- Deaths: 74.7% White (570 White / 763 race known, 787 race unknown, 50.0% of 1550 total)

**FA State COVID-19 Mortality/Death rates from 2020-04-26**:
- 4.2% Black (166 Black deaths / 4027 Black cases) 6.9% White (570 White deaths / 8292 White cases) 5.9% All race known (763 race known deaths / 12915 race known cases) 2.6% All race unknown (787 race unknown deaths / 28250 race unknown cases) 3.6% All (1550 all deaths / 41165 all cases)

**ACHD COVID-19** data from 2020-04-26:
- All Cases: 23.2% Black (210 Black / 905 race known, 306 race unknown, 25.3% of 1211 total)
- All Tests: 21.4% Black (1549 Black / 7251 race known, 8340 race unknown, 53.5% of 15591 total)
- Hospitalizations: 24.0% Black (48 Black / 200 race known, 13 race unknown, 6.1% of 213 total)
- Deaths: 16.7% Black (9 Black / 54 race known, 19 race unknown, 26.0% of 73 total)

**ACHD COVID-19 Mortality/Death rates from 2020-04-26**:
- 4.3% Black (9 Black deaths / 210 Black cases) 6.7% White (45 White deaths / 675 White cases) 6.0% All race known (54 race known deaths / 905 race known cases) 6.2% All race unknown (19 race unknown deaths / 306 race unknown cases) 6.0% All (73 all deaths / 1211 all cases)
Outcome: Open data!

- Co-designed open data feeds with Allegheny County to inform actions to eliminate disparity in COVID response.

- Provided context about data systems in a guide for users enabling more-responsible uses of this data.

Outcome: BEC COVID-19 Data Dashboard visualizes the pandemic’s impact on the Black community

https://covid.createlab.org/
Goal: Use vaccination data to track progress and inform outreach

- Collected and shared potential use cases for vaccine data with the state once we learned that they were building data dashboards without community feedback.
“Many of the challenges that have been exposed by the pandemic can best be addressed together with people who are marginalized, and we’re asking that our public-sector partners climb the ladder of participation and work in meaningful partnership with us to institutionalize the values of equity and justice in our public health infrastructure.”
There are very large differences in Black vaccination rates between communities.

<table>
<thead>
<tr>
<th>Community</th>
<th>Fully-Vaccinated (age 5 +)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monroeville</td>
<td>73%</td>
</tr>
<tr>
<td>Crawford-Roberts</td>
<td>62%</td>
</tr>
<tr>
<td>Stanton Heights</td>
<td>60%</td>
</tr>
<tr>
<td>Penn Hills</td>
<td>52%</td>
</tr>
<tr>
<td>Swissvale</td>
<td>45%</td>
</tr>
<tr>
<td>Wilkinsburg</td>
<td>45%</td>
</tr>
<tr>
<td>East Hills</td>
<td>38%</td>
</tr>
<tr>
<td>Lincoln-Lemington-Belmar</td>
<td>37%</td>
</tr>
<tr>
<td>McKeesport</td>
<td>35%</td>
</tr>
<tr>
<td>Homewood North</td>
<td>33%</td>
</tr>
<tr>
<td>Duquesne</td>
<td>32%</td>
</tr>
<tr>
<td>Knoxville</td>
<td>31%</td>
</tr>
<tr>
<td>Marshall-Shadeland</td>
<td>28%</td>
</tr>
<tr>
<td>Northview Heights</td>
<td>25%</td>
</tr>
<tr>
<td>Rankin</td>
<td>25%</td>
</tr>
</tbody>
</table>

ACHD as of 1/22/2022 - selected communities with over 1K Black People
PANEL QUESTIONS
Recommendations for *federal agencies*:

- Enhance federal infrastructure and engage communities as part of the process to prepare data systems for the next emergency.
- Provide information about context, biases, and limitations about data in the form of guides/readme's and datasheets.
- Build relationships with local coalitions and national communities of practice (such as National Neighborhood Indicators Partnership).
- Improve and enhance data sharing among local community partners, local/state health departments, health plans, etc.
- Provide federal funds to support community data partnerships, especially at the local level.
Recommendations for *national foundations*:

- Provide incentives for coalition and relationship development in advance of the next crisis.

- Provide foundation funds to support community data partnerships, especially at the local level.

- Support local participation in national communities of practice. Our local efforts have been informed by the National Neighborhood Indicators Partnership, Actionable Intelligence for Social Policy, Data Across Sectors for Health, and The Data Foundation, among other communities.
Thank you!

www.blackequitypgh.org

info@blackequitypgh.org
5-minute Break

Event will resume at 1:55 p.m. ET
A dynamic, inclusive approach to learning agenda development for the Centers for Disease Control and Prevention’s (CDC’s) Center for State, Tribal, Local, and Territorial Support (CSTLTS): Reflections on the participant engagement process

Elizabeth Douglas, Senior Manager, ICF / @ICF

Jessie Rouder, Lead Research Scientist, Behavioral Health, ICF / @ICF
A Dynamic, Inclusive Approach to Learning Agenda Development

For the Centers for Disease Control and Prevention’s (CDC’s) Center for State, Tribal, Local, and Territorial Support (CSTLTS)

Reflections on the Participant Engagement Process

June 23, 2022
Presenters

Elizabeth Douglas, MS
Senior Manager, Research Science
ICF
Atlanta, GA

Jessie Rouder, MA
Lead Research Scientist, Data Visualization and Storytelling
ICF
Long Island, New York
What is a learning agenda?
Why is CSTLTS creating a learning agenda?
What is the approach and process?
What are the lessons learned and key takeaways?
What is a learning agenda?

A learning agenda is comprised of a set of prioritized questions about evidence needs to inform future decision-making in an organization. The questions and analytical approaches to address the questions are collaboratively developed by organizational leaders, staff, and stakeholders.¹

An inclusively- and strategically- developed learning agenda provides a list of important questions as well as plans for addressing the questions, balancing the interests, informational needs, and time horizons for different organizational decision-makers.²
CDC’s Center for State, Tribal, Local and Territorial Support (CSTLTS) Science Unit and CSTLTS Learning Agenda Steering Committee

CSTLTS is CDC’s primary connection to health officials and leaders of state, tribal, local, and territorial (STLT) public health agencies as well as other government leaders who work with health departments.

CSTLTS improves community health outcomes by strengthening STLT public health agencies.

ICF

ICF is a global consulting services company whose work helps federal agencies and organizations solve their most complex challenges. Since 1969, public and private sector clients have worked with ICF to navigate change and shape the future.
CSTLTS Strategic Map Priorities

**PRIORITY 1.** Strategic partnership and stakeholder engagement

**PRIORITY 2.** Elite public health workforce

**PRIORITY 3.** Operational excellence of health departments

**PRIORITY 4.** Access to data and evidence about and for STLT health departments

**PRIORITY 5.** Equity, diversity and inclusion

WHY IS CSTLTS CREATING A LEARNING AGENDA?

• Help CSTLTS tell the story of their services and outcomes and inform improvement

• Address topics of importance in the CSTLTS Strategic Map by focusing on priority questions and revisit them every 3 years for relevancy

• Strengthen the evidence base for CSTLTS programs and services

• Assist in driving efficient use of resources
WHAT IS THE APPROACH?

• Identify and engage internal agency staff and external partners

• Build a shared understanding of the prioritization process and outcomes

• Collaborate to generate and prioritize the learning agenda questions

• Design the evidence building activities to address the final prioritized learning agenda questions
WHO WAS ENGAGED?

CSTLTS subject matter experts and leaders 35
CDC thought leaders
STLT public health organization representatives 33
STLT health department representatives

CDC staff
External representatives
External Participants

**STLT Health Departments**
- Alaska Department of Health and Social Services
- Arkansas Department of Health
- California Department of Public Health
- Cherokee Nation Health Services
- Columbus Public Health
- Detroit Health Department
- Granville-Vance District Health Department
- Idaho North Central District
- Kentucky Department for Public Health
- Mecklenburg County Health Department
- New Mexico Department of Health
- Pennsylvania Department of Health
- San Antonio Metropolitan Health District
- Tulsa Health Department
- Winnebago Public Health Department

**Public Health Organizations**
- American Public Health Association
- Association of State and Territorial Health Officials
- CDC Foundation
- ChangeLab Solutions
- deBeaumont Foundation
- Indiana University Richard M. Fairbanks School of Public Health
- Indigenous Wellness Research Institute at the University of Washington
- Mississippi Public Health Association
- National Association of County and City Health Officials
- National Network of Public Health Institutes
- Pacific Island Health Officers’ Association
- Public Health Accreditation Board
- Public Health Foundation
- Robert Wood Johnson Foundation
<table>
<thead>
<tr>
<th>Generate</th>
<th>Refine</th>
<th>Prioritize</th>
<th>Finalize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question-generating Workshops</td>
<td>Discussions with Thought Leaders</td>
<td>Delphi Panel Survey</td>
<td>Finalized Learning Agenda Questions</td>
</tr>
</tbody>
</table>

- **Generate**
  - Question-generating Workshops

- **Refine**
  - Discussions with Thought Leaders

- **Prioritize**
  - Delphi Panel Survey

- **Finalize**
  - Finalized Learning Agenda Questions
Question-generating Workshops

Discussions with Thought Leaders

Delphi Panel Survey

Finalized Learning Agenda Questions
Process for Generating and Prioritizing Learning Agenda Questions

1. **Orientation Webinar**
   Aim: Ensure participants are knowledgeable about the learning agenda context, key terms, and process.

2. **Question-generating Workshops**
   Aim: Facilitate virtual collaboration, ideation, and draft question generation.

3. **Question Review and Refinement**
   Aim: Review and refine draft learning agenda questions asynchronously.

4. **Discussions with CDC Thought Leaders**
   Aim: Consult thought leaders to assess question alignment with other agency priorities.

5. **Delphi Panel**
   Aim: Review, reassess, and rate the draft learning agenda questions through multiple rounds of an asynchronous survey.
Lessons Learned and Takeaways

• Invest time in developing a shared understanding among participants

• Emphasize defining the nature of a learning agenda question and related parameters

• Plan participant activities well in advance to align schedules, ensure availability, and promote engagement

• Integrate multiple engagement methods and time points to maximize participant engagement

• Remain flexible and innovative throughout the process

• Consider an approach that works best given your agency’s context (e.g., a phased, iterative approach for CSTLTS)
Thank You and Contact Information

• Elizabeth Douglas, Elizabeth.Douglas@icf.com
• Jessie Rouder, Jessie.Rouder@icf.com

• Special thanks to the CSTLTS Science Unit Team
  - Andrea Young, Associate Director for Science
  - Tamara Lamia, Senior Health Scientist
  - Stephanie Koh, Health Scientist

The content of this presentation are those of the authors and do not necessarily represent the official position of or endorsement by the Centers for Disease Control and Prevention.
References


3. CSTLTS Strategic Map: https://www.cdc.gov/publichealthgateway/strategy/strategic-map.html


Best Practices for Monitoring and Evaluating the ARP, IIJA and Other Programs: Report of the Department of Commerce Data Governance Working Group

Ron Jarmin, Deputy Director, U.S. Census Bureau / @jarmin_ron

Whitney Duffey Jones, Senior Advisor & Chief of Staff to the CFO, U.S. Census Bureau / @uscensusbureau

Carla Medalia, Assistant Division Chief for Business Development, Economic Reimbursable Surveys Division, U.S. Census Bureau / @uscensusbureau

Ben Page, Chief Financial Officer, U.S. Census Bureau / @uscensusbureau

Ryan Smith, Policy Advisor for the Office of Regional Affairs, Economic Development Administration / @US_EDA

Oliver Wise, Chief Data Officer, U.S. Department of Commerce / @ojwise
Best practices for monitoring and evaluating the ARP, IIJA and other programs:

Report of the Department of Commerce Data Governance Working Group

June 23, 2022

Data Foundation Virtual Symposium

Ben Page
Agenda

1. Executive summary, roadmap, and recommendations
2. Census Bureau’s Data Linkage Infrastructure for program evaluation
3. Data needs from an implementing bureau’s perspective
4. Next steps from Chief Data Officer’s perspective
5. Discussion
Panelists

• Ron Jarmin, Deputy Director, U.S. Census Bureau
• Whitney Duffey Jones, Senior Advisor & Chief of Staff to the CFO, U.S. Census Bureau
• Carla Medalia, Assistant Division Chief for Business Development, Economic Reimbursable Surveys Division, U.S. Census Bureau
• Ben Page, Chief Financial Officer, U.S. Census Bureau
• Ryan Smith, Policy Advisor for the Office of Regional Affairs, Economic Development Administration
• Oliver Wise, Chief Data Officer, U.S. Department of Commerce
Executive summary, roadmap, and recommendations

June 23, 2022
Data Foundation Virtual Symposium
Ron Jarmin
A Once-in-a-Generation Opportunity

American Rescue Plan (ARP) and Infrastructure Investment and Jobs Act (IIJA) provide unprecedented resources to improve America’s infrastructure and support economic resilience and long-term growth.

We need to:

- Ensure we’re using taxpayer dollars wisely and effectively
- Quickly and accurately aggregate data on program performance
- Leverage data as a strategic asset
- Enable evidence-based decision making
- Advance data sharing and collaboration opportunities
- Expand model beyond DOC and to future programs
Department of Commerce
Data Governance Working Group (DGWG)

Purpose:
• To identify ways to collect incoming program performance information
• Minimize burden to aid recipients
• Maximize ability to report achievement of program goals and objectives
• Demonstrate lessons learned
• Improve program outcomes
• Foster adoption of promising practices
Participants

Members:
- DOC Implementation Coordination Office
- U.S. Census Bureau (chair)
- Economic Development Administration (EDA)
- National Oceanic and Atmospheric Administration (NOAA)
- National Telecommunications and Information Administration (NTIA)
- National Institute of Standards and Technology (NIST)
- Bureau of Economic Analysis (BEA)
- DOC Office of the Under Secretary of Economic Affairs (including Chief Data Officer and Evaluation Officer)

Advisors:
- Office of Management and Budget (OMB)
- Department of Transportation (DOT)
- Staff from the DGWG member agencies
Scope and phased deliverables

• **Phase 1** (completed)
  • Develop shared data structure and data quality standards to facilitate data linkages and reduce reformatting complications
  • Describe strategies to ensure availability of high-quality data to support policy and program outcomes

• **Phase 2** (starting soon)
  • Identify common metadata standards to ensure the DOC’s data can be leveraged as a strategic asset

• **Phase 3** (future - planned)
  • Discuss strategies to address barriers to data collection/use
  • Implement systems standards to ensure maximum interoperability at an enterprise level
DGWG Report

• “Best Practices for Monitoring and Evaluating the ARP, IIJA and other programs: Report of the Department of Commerce Data Governance Working Group”

• Report is the culmination of “phase 1” sprint, and:
  • Establishes goal for program monitoring and evaluation
  • Compares evaluations strategies
  • Identifies data resources and linkage strategies
  • Considers factors such as geographies, socioeconomic factors, equitable delivery, and environmental risk factors
Recommendations

• Agencies should follow all applicable standards issued by the Office of Management and Budget (OMB) for grant awardees to report on geographic location of primary and secondary awardees, as well as the point of service delivery. The DGWG supports using census tracts as a standard reporting element.

• Agencies should continue to measure and monitor program operations (e.g., timeliness, compliance with regulations) following or exceeding guidance from OMB and their respective agencies.

• Agencies should leverage existing impact projection models from industry, academia, or government where possible.
Recommendations

• Agencies should implement a measurement and evaluation design that emphasizes credible results within resource, cost, and schedule constraints.
• Agencies should consider implementing large-scale observational studies that link program administrative data to previously collected data from censuses, surveys, administrative records, commercial vendors, and aggregated indices. The Census Bureau’s Data Linkage Infrastructure may be particularly useful.
• Agencies should collect high-quality unique identifiers from aid awardees, including both primary and secondary awardees to enable linkage to other data sources.
Recommendations

• Agencies should evaluate programs based on the phase of program implementation, direct versus indirect program impacts, and projected versus observed program impacts

• Phases
  • Program Design/Stand Up
  • Pre-Award Program Implementation
  • Post-Award Program Implementation
  • Closeout
Recommendations

• Agencies should use metrics and available indices of economic and geographic vulnerability to determine the equity or bias in program delivery.

• Agencies should consider incorporating program impacts on the environment and climate resilience into their evaluation plans.

• Agencies should establish or participate in a working group to identify existing standards and best practices for managing program operations and evaluation data and develop guidance for use by data practitioners.

• Agencies should create a community of practice to share lessons learned from program implementation and foster collaboration.
Framework for program evaluation

<table>
<thead>
<tr>
<th>Phase</th>
<th>Direct program impacts (D)</th>
<th>Indirect program impacts (I)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Projected</td>
<td>Projected</td>
</tr>
<tr>
<td>Program design/stand up</td>
<td>Identify direct impacts ($D$) and determine how to capture in NOFOs</td>
<td>Identify indirect impacts ($I$) and determine administrative/survey data to measure</td>
</tr>
<tr>
<td>Pre-award program implementation</td>
<td>What is potential benefit of $D$?</td>
<td>What is potential benefit of $I$?</td>
</tr>
<tr>
<td></td>
<td>What is current (baseline) rate of $D$?</td>
<td>Baseline: What is current rate of $I$?</td>
</tr>
<tr>
<td>Post-award program implementation</td>
<td>Compare to initial projection of $D$; update projection of $D$ based on observed</td>
<td>Compare to initial projection of $I$; update projection of $I$ based on observed</td>
</tr>
<tr>
<td></td>
<td>Compare to baseline levels of $D$</td>
<td>Compare to baseline levels of $I$</td>
</tr>
<tr>
<td>Closeout</td>
<td>Compare to projected levels of $D$ (from each phase)</td>
<td>Compare to projected levels of $I$ (from each phase)</td>
</tr>
<tr>
<td></td>
<td>Compare to baseline levels of $D$ and implementation levels of $D$</td>
<td>Compare to baseline levels of $I$ and implementation levels of $I$</td>
</tr>
</tbody>
</table>
Census Bureau’s Data Linkage Infrastructure for program evaluation

June 23, 2022
Data Foundation Virtual Symposium
Carla Medalia
carla.medalia@census.gov
DGWG report: leverage existing survey and administrative data

- Agencies should leverage existing survey and administrative data whenever possible when developing and executing program evaluation
  - Particularly useful to enable large scale observational studies
  - Enables consistent program evaluation across agencies and facilitates comparisons across variety of programs
  - Reduces burden of aid awardees when reporting on the work they did with the federal funds
  - Allows agencies to answer questions otherwise not possible to answer
Support high-quality research and evaluation, advancing the Census Bureau's mission of providing timely and unbiased data to support evidence-based decision making.
Census data: people and households

Sources
• Decennial censuses
• American Community Survey
• Current Population Survey
• Survey of Income and Program Participation
• American Housing Survey

Uses
• Sociodemographic characteristics
• Income, poverty, health insurance
• Labor force, occupation, industries
• Data linked over time
• Individuals, families, households
• Geographic information down to Census blocks
Census data: businesses, governments, and economy

Sources
• Economic Censuses
• Census of Governments
• Firm Surveys
• Establishment Surveys
• Transaction or Trade data
• Longitudinal Employer-Household Dynamics (LEHD)

Uses
• Microdata about all U.S. businesses: precise geolocations
• Microdata about large samples of U.S. business: geolocations, payroll, tax records, foreign investments
• Detailed geographies and industries
• Data linked over time
• Employee and employer linked data
Administrative data

**Federal data**
- Bureau of Labor Statistics
- Bureau of Prisons
- Corporation for National and Community Service
- Department of Defense
- Department of Veterans Affairs
- Health and Human Services
- Housing and Urban Development
- Indian Health Service
- Internal Revenue Service
- Office of Personnel Management
- Selective Service System
- Small Business Administration
- Social Security Administration
- U.S. Postal Service

**State/local data**
- Unemployment Insurance
- Supplemental Nutrition Assistance Program
- Women, Infants, and Children
- Temporary Assistance for Needy Families
- Low Income Energy Assistance Program
- Child Care Subsidy
- Homeless Management Information System
- Alaska Permanent Fund
- Puerto Rico tax data
- California tax data
- University education data
- Criminal Justice Administrative Records System

**Third party/commercial data**
- Mortgage/address data
- Contact frame
- Property and tax foreclosure

Note: data access subject to approval
Data linkage quality depends on PII/BII

<table>
<thead>
<tr>
<th>Linkage type</th>
<th>Linkage method</th>
<th>Linkage fields to collect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address linkage</td>
<td>Master Address File Match (MAFMATCH)</td>
<td>Full address; coordinates for point of service delivery</td>
</tr>
<tr>
<td>Organization</td>
<td>TF-IDF (term frequency-inverse document frequency), MAMBA</td>
<td>Businesses/non-profits: For both establishment and firm collect: EIN, business name, mailing address, physical location address, NAICS, company web address, UEI, SSN (owner of sole proprietorship) Governments: name, address, UEI, web address</td>
</tr>
<tr>
<td>Person linkage</td>
<td>Person Identification Validation System (PVS)</td>
<td>Full name (first, middle, last, suffix), complete date of birth (age is acceptable but less optimal), full address, sex, SSN/ITIN (for administrative records with authority to collect)</td>
</tr>
</tbody>
</table>
Legal framework and data governance

• Legal Framework: U.S.C. Title 13
  • § 6. Acquire and utilize records to the greatest extent possible
  • § 8. Reimbursable studies and joint statistical projects
  • § 9. Protect confidential individual and establishment data, limit access, and statistical uses
  • § 23(c). Swear in researchers to assist the Census Bureau
  • § 214. Wrongful disclosure of information

• Data governance infrastructure
  • Data stewardship, disclosure review
  • Anonymized data
  • Secure computing environments
Census Bureau’s Data Linkage Infrastructure enables measurement of program impacts

• Direct program impacts: defined in legislation
• Indirect program impacts: second-order effects
• Example: NTIA’s Broadband Equity, Access, and Deployment (BEAD) Program
  • Direct: broadband built
  • Indirect: access to internet; educational attainment; commuting patterns
Prototype library of indirect program impact metrics

<table>
<thead>
<tr>
<th>Category</th>
<th>Measure</th>
<th>Data source name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>Population characteristics</td>
<td>American Community Survey, decennial census</td>
</tr>
<tr>
<td>Demographic/ business</td>
<td>Business and owner characteristics</td>
<td>Annual Business Survey, Nonemployer Statistics by Demographics</td>
</tr>
<tr>
<td>Economic indicators</td>
<td>New business starts</td>
<td>Business Formation Statistics and Business Dynamics Statistics</td>
</tr>
<tr>
<td>Economic indicators</td>
<td>Manufacturing</td>
<td>Manufacturers’ Shipments, Inventories, and Orders (M3)</td>
</tr>
<tr>
<td>Environment/ Climate</td>
<td>Coastal communities and businesses</td>
<td>American Community Survey; Annual Business Survey; Business Register</td>
</tr>
<tr>
<td>Jobs</td>
<td>Labor force participation rate</td>
<td>Current Population Survey, American Community Survey</td>
</tr>
<tr>
<td>Other economic</td>
<td>Tax revenues</td>
<td>Census of Governments</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>Poverty Rate</td>
<td>CPS ASEC, ACS, administrative data</td>
</tr>
<tr>
<td>Underserved communities index</td>
<td>At-risk neighborhoods</td>
<td>Community Resilience Estimates</td>
</tr>
</tbody>
</table>
US Economic Development Administration – “Data Needs from an Implementing Bureau’s Perspective”

- **Ryan Smith** – Policy Advisor for the Office of Regional Affairs and Data Evolution Lead for EDA
- EDA is seeking candidates for a Chief Data Officer position ([rsmith2@eda.gov](mailto:rsmith2@eda.gov) & [EDACareer@eda.gov](mailto:EDACareer@eda.gov))
- [https://eda.gov/careers/opportunities/](https://eda.gov/careers/opportunities/)

EDA is seeking to hire a CDO.
US Economic Development Administration

**EDA 101**

- US EDA is a bureau within the US Dept. of Commerce
- Mission: “To lead the federal economic development agenda by promoting innovation and competitiveness, preparing American regions for growth and success in the worldwide economy.”
- Investment Priorities: [https://eda.gov/about/investment-priorities/](https://eda.gov/about/investment-priorities/)
- Core Business Lines –
  - Grant Making and Capacity Building in Economically Distressed Areas (Public Works, EAA)
  - Build-to-Scale investments supporting Innovation and Entrepreneurship
  - Disaster Recovery via the Economic Recovery Support Functions and Disaster Supplemental

EDA is seeking to hire a CDO.
Government Performance and Results Act – 
*GPRA 101*

- EDA typically measures outcomes based on Jobs Created or Retained and Private Investment Dollars Leveraged
- Supported by EDA’s Logic Model
- EDA Geographic Eligibility is dependent on Economic Distress
- Investment Priorities – In Particular Equity/Underserved
- Reporting happens at 3-6-9 years after award, Validated and Verified

EDA is seeking to hire a CDO.
## Prototype library of indirect program impact metrics

<table>
<thead>
<tr>
<th>Category</th>
<th>Measure</th>
<th>Data source name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>Population characteristics</td>
<td>American Community Survey, decennial census</td>
</tr>
<tr>
<td>Demographic/business</td>
<td>Business and owner characteristics</td>
<td>Annual Business Survey, Nonemployer Statistics by Demographics</td>
</tr>
<tr>
<td>Economic indicators</td>
<td>New business starts</td>
<td>Business Formation Statistics and Business Dynamics Statistics</td>
</tr>
<tr>
<td>Economic indicators</td>
<td>Manufacturing</td>
<td>Manufacturers’ Shipments, Inventories, and Orders (M3)</td>
</tr>
<tr>
<td>Environment/Climate</td>
<td>Coastal communities and businesses</td>
<td>American Community Survey; Annual Business Survey; Business Register</td>
</tr>
<tr>
<td>Jobs</td>
<td>Labor force participation rate</td>
<td>Current Population Survey, American Community Survey</td>
</tr>
<tr>
<td>Other economic</td>
<td>Tax revenues</td>
<td>Census of Governments</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>Poverty Rate</td>
<td>CPS ASEC, ACS, administrative data</td>
</tr>
<tr>
<td>Underserved communities index</td>
<td>At-risk neighborhoods</td>
<td>Community Resilience Estimates</td>
</tr>
</tbody>
</table>
How We Hope to Use Data & Linkages

- **Evidence Building** for Program Effectiveness and Future Appropriations
- Inputs for Other **Projects** and Public Data **Products** (NERDE) – Understanding Eligibility and Impact
- Tying administrative data via DOC-wide investments in IT systems/infrastructure will connect EDA’s mission space and project information with other DOC equities
- **EDA is recruiting for and seeking to hire a Chief Data Officer**

- **Ryan Smith** - (rsmith2@eda.gov)
  
  *Policy Advisor for the Office of Regional Affairs and Data Evolution Lead for EDA*
Next steps from Chief Data Officer’s Perspective

Oliver Wise, Chief Data Officer, Department of Commerce
Discussion
Data Foundation Supporters

Support our work by donating at datafoundation.org
A special thank you to our partner

Trachtenberg School of Public Policy & Public Administration

THE GEORGE WASHINGTON UNIVERSITY
Thank you for tuning in to Day 2 of our Virtual Research Symposium!

Learn more at datafoundation.org, or follow us on social media @data_foundation