

Connecting families to benefits using administrative data: a case study in hashed linkages

Evan White
Executive Director
California Policy Lab at UC Berkeley

evanbwhite@berkeley.edu

October 2022

FTA Revenue Estimation and Tax Research Conference



Joint work with...



STATE OF CALIFORNIA
Franchise Tax Board



CODE *for*
AMERICA



The California Policy Lab



- Research institute at Univ of California, that works almost exclusively with person-level government 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
 - Criminal arrests, charges, and dispositions
 - K12 and higher-ed enrollments and outcomes
- Research on...
 - Social Safety Net
 - Labor / Employment
 - Homelessness
 - Criminal Justice
 - Education
 - Health

Anti-poverty benefits, delivered via tax system



- Anti-poverty tax credits: cash delivered through tax filing with potential to lift millions out of poverty
 - Federal Earned Income Tax Credit
 - Federal stimulus payments
 - Advanced Child Tax Credit (2020-21)
 - Plus state versions of all of these (in CA, all three)

This is a lot of money



**One child under age 5,
Earnings: \$5,000**

**Total
credits**

EITC

\$1,692

CalEITC

\$1,438

YCTC

\$1,000

Golden state stimulus

\$600

Advanced CTC

\$1,800

EIP 1, 2, 3 adult

\$3,200

EIP 1, 2, 3 child

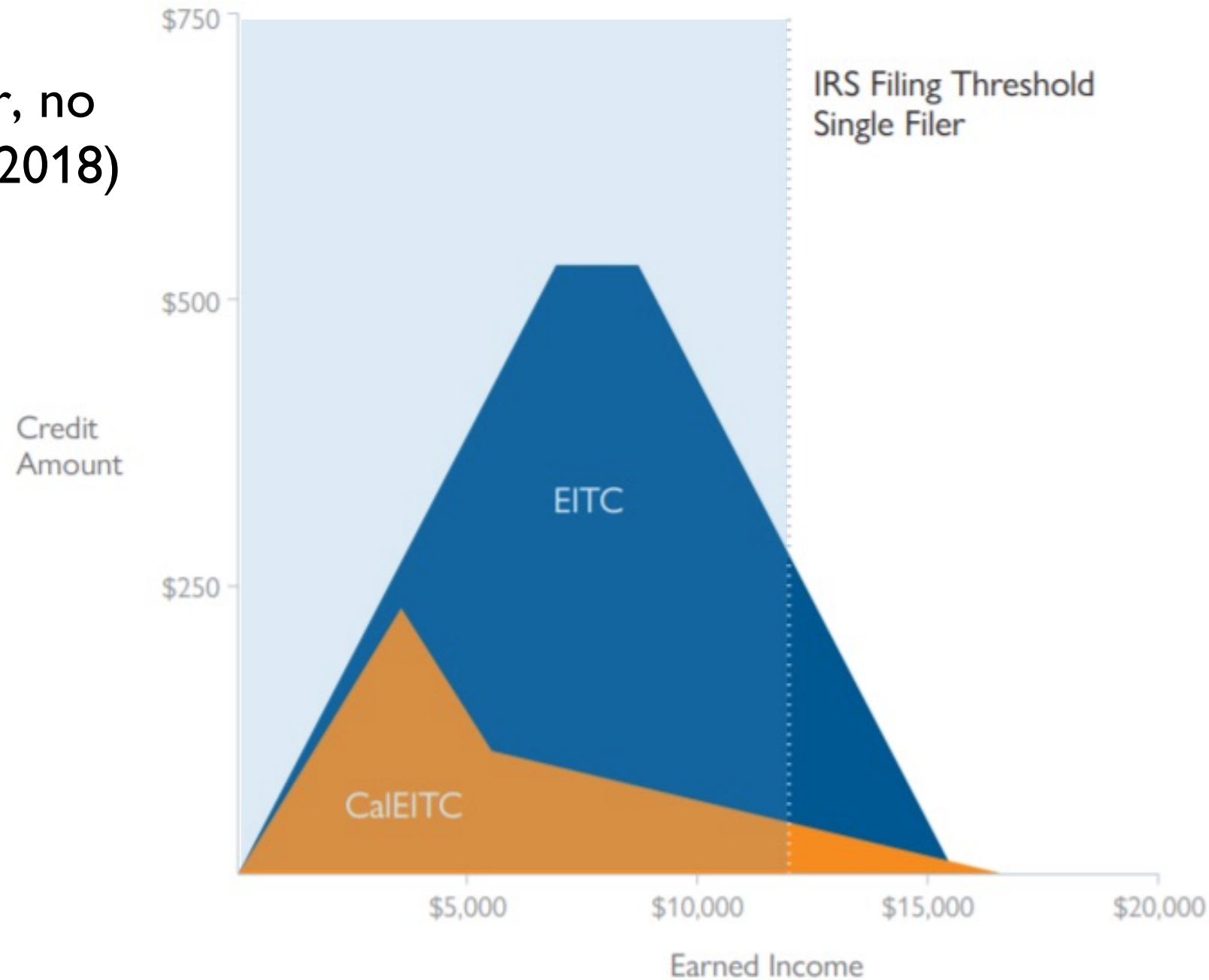
\$2,500

Total

\$12,230 (over 2 years)

Many who qualify fall below the filing threshold

Single filer, no
children (2018)





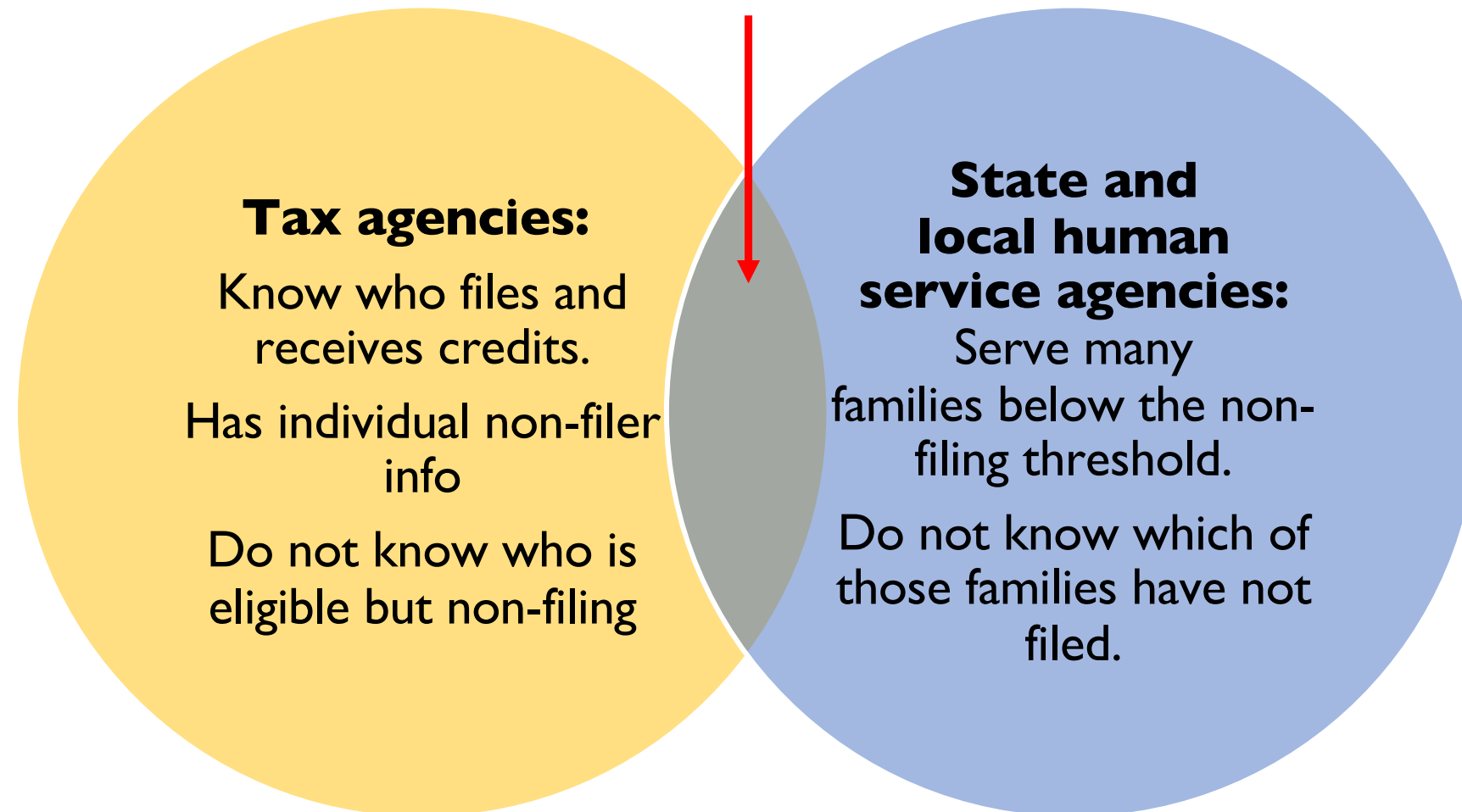
Three vexing policy questions

- 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 has all the data



Take-up: $\frac{\text{Number of eligible non-filers}}{\text{All eligible individuals}}$

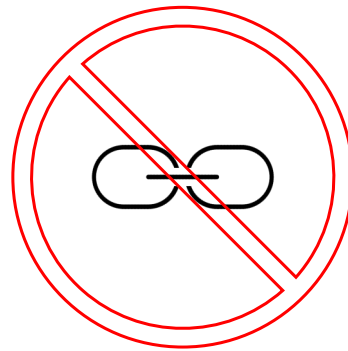


The privacy challenge



Agency 1

"I don't want to
share my PII"



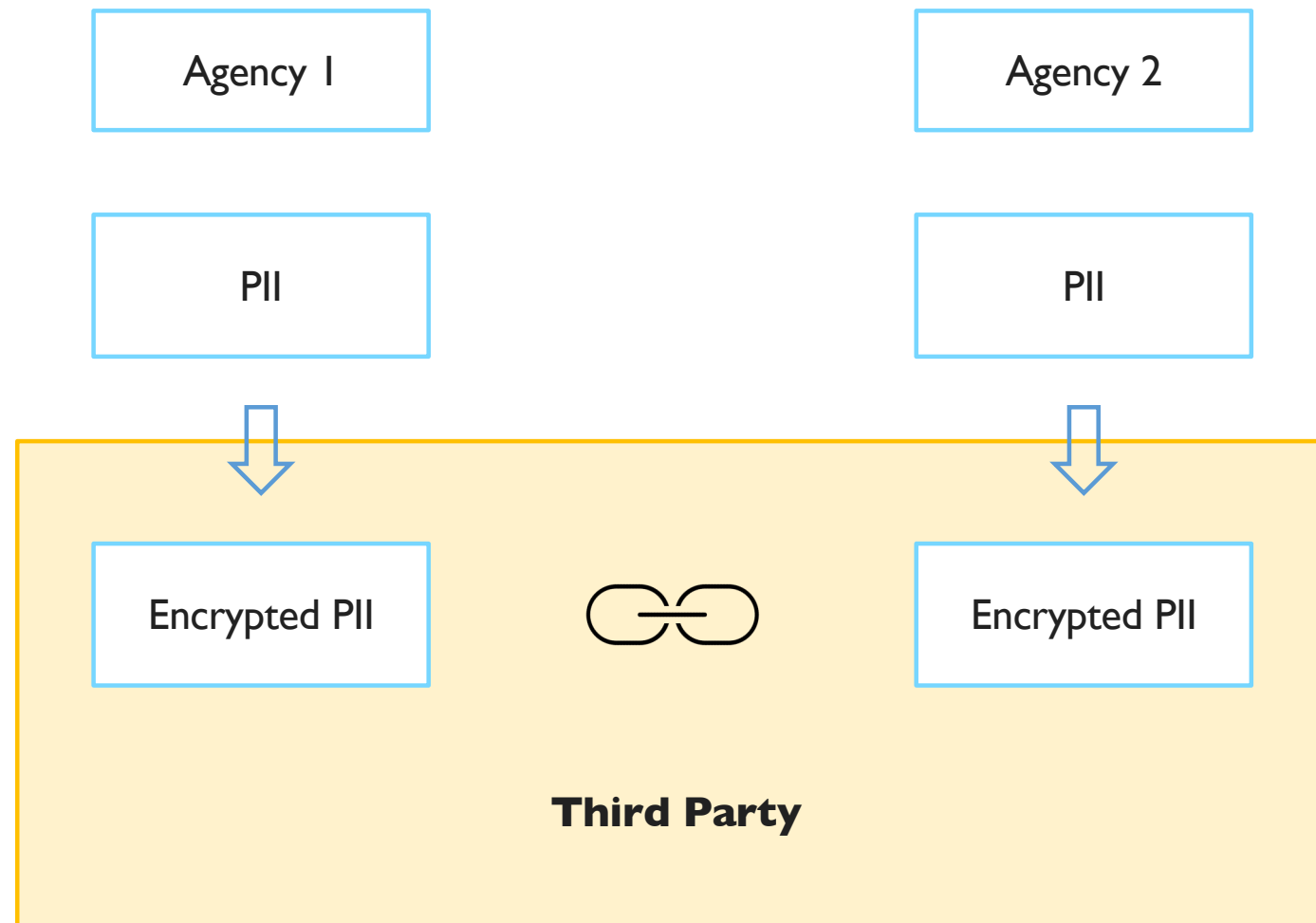
Agency 2

"I don't want to
share my PII"

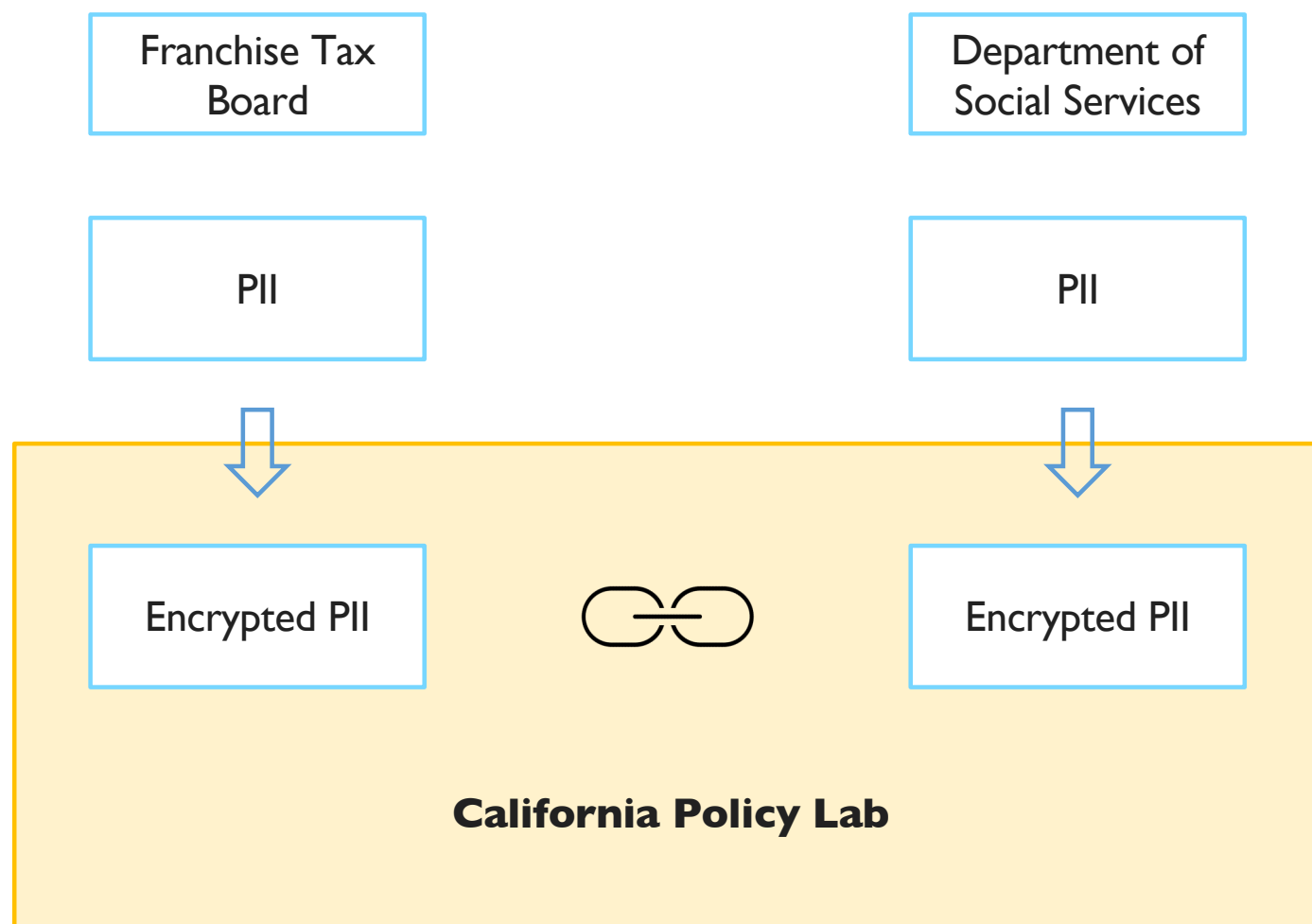
Rationales

- 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

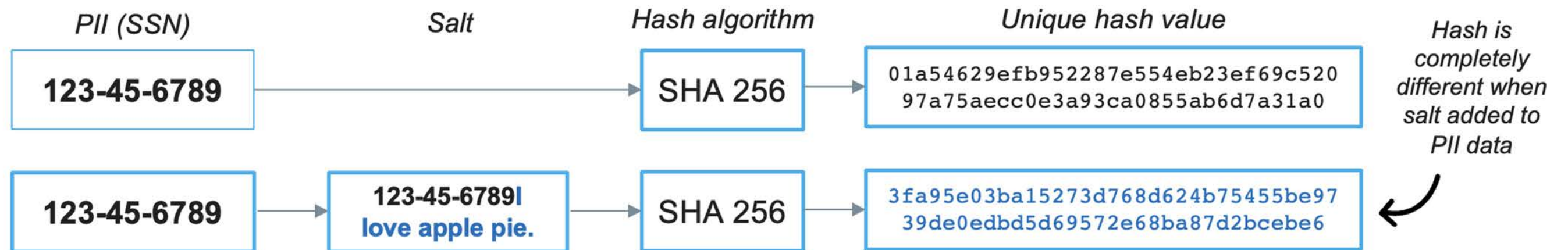
- One-directional encryption
- Mathematically impossible to reverse
- Hash SSN, first name, last name, and date of birth



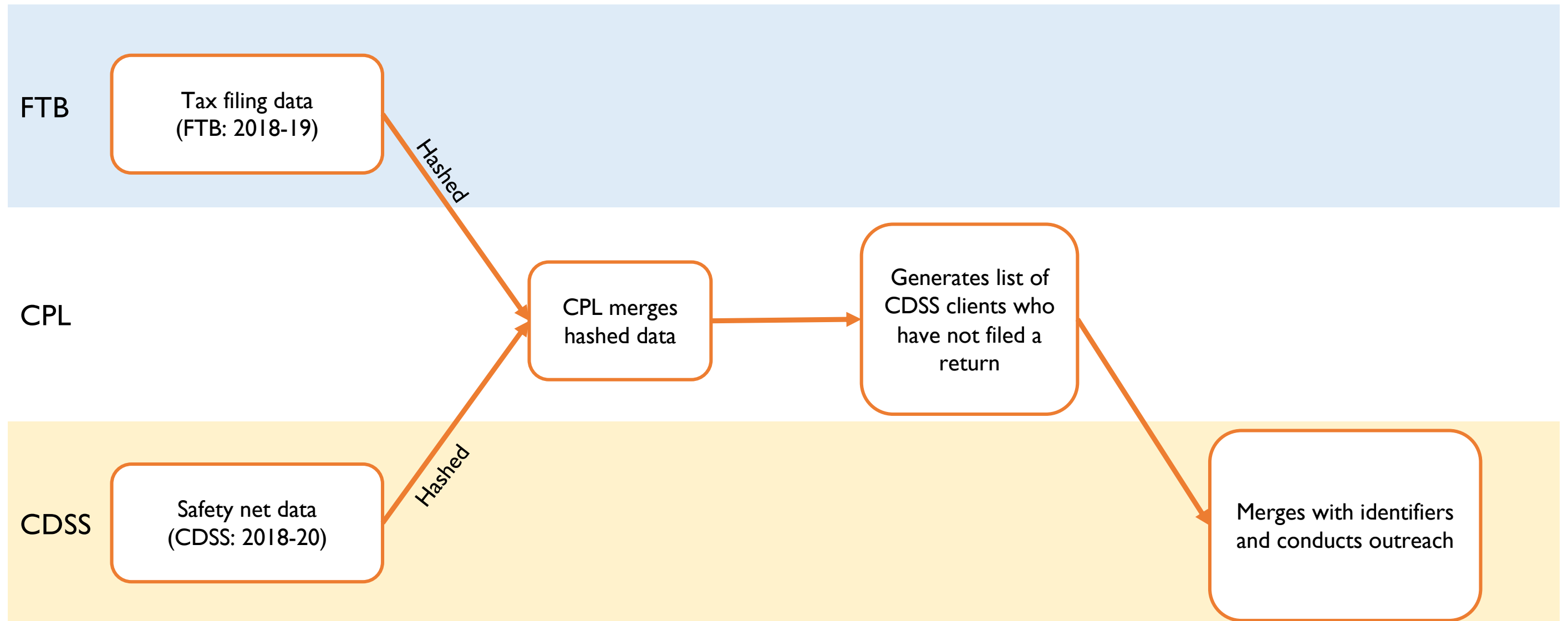
Hashing and the “salt”



- Mitigates risks from guessing inputs
- Pick something strong



Process flow



**1st result: the tax system does a good job for
75%**



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

75%

received the federal stimulus automatically

74%

of children received child tax credit automatically

2nd result: single adults with very low earnings fare worst



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



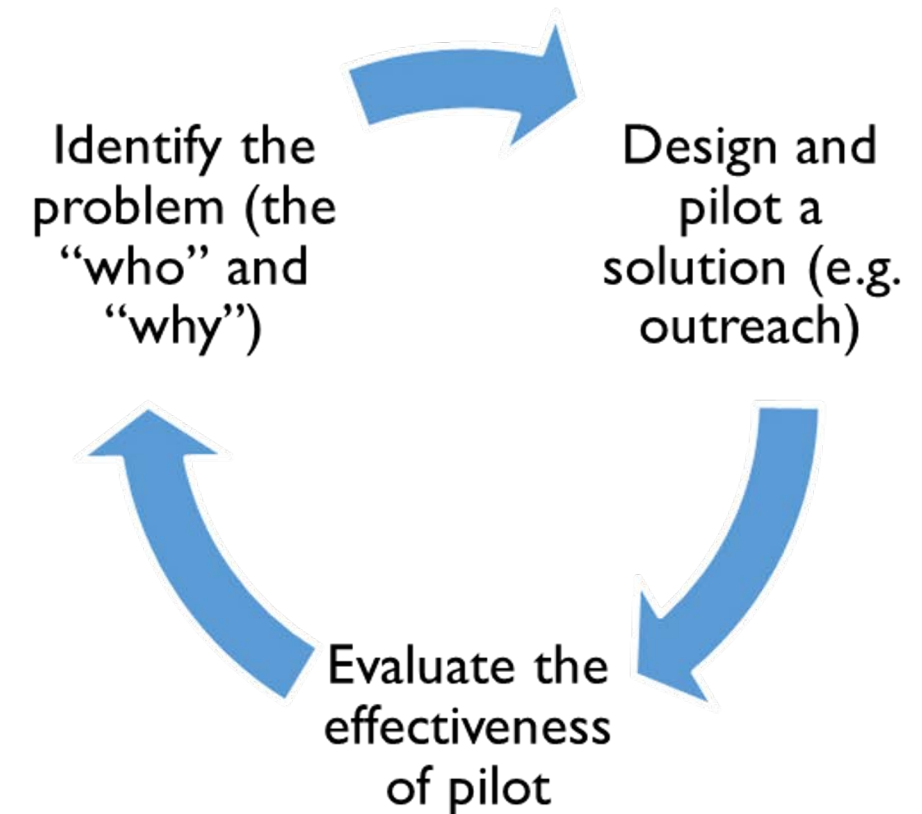
Three vexing policy questions

- 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?

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
- CDSS 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 returners?

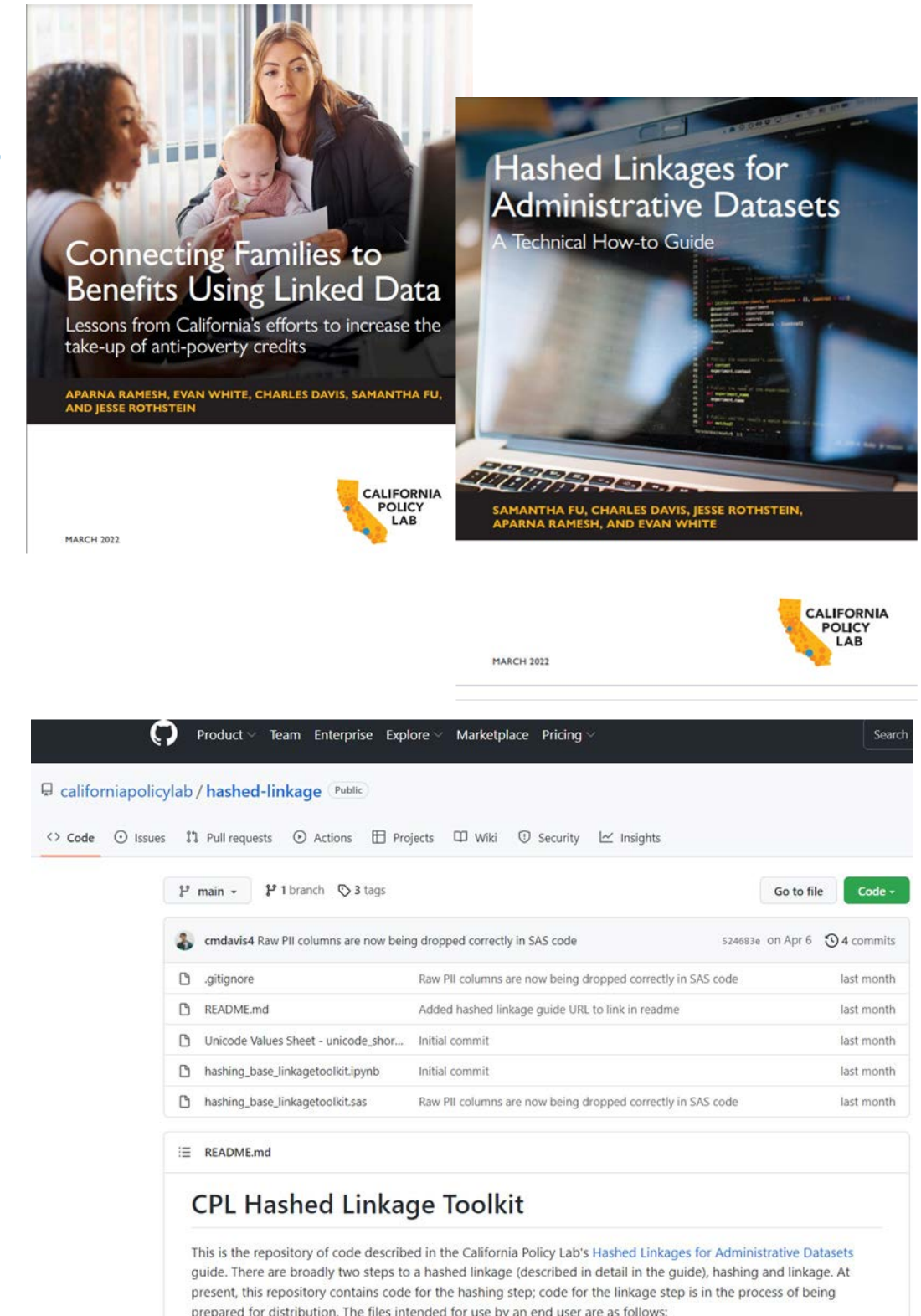


Want to administer hashed linkage?

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/californiapolicylab/hashed-linkage>



Steps to defining and executing a linkage

Determine the linkage strategy

1

Define your goal

2

Identify barriers
to linkage

3

Decide on best
pathway to
linkage

4

Decide whether
you'll need a
third party

Implement a hashed linkage

5

Budgeting for
time and
resources

6

Set up legal
framework and
agreements

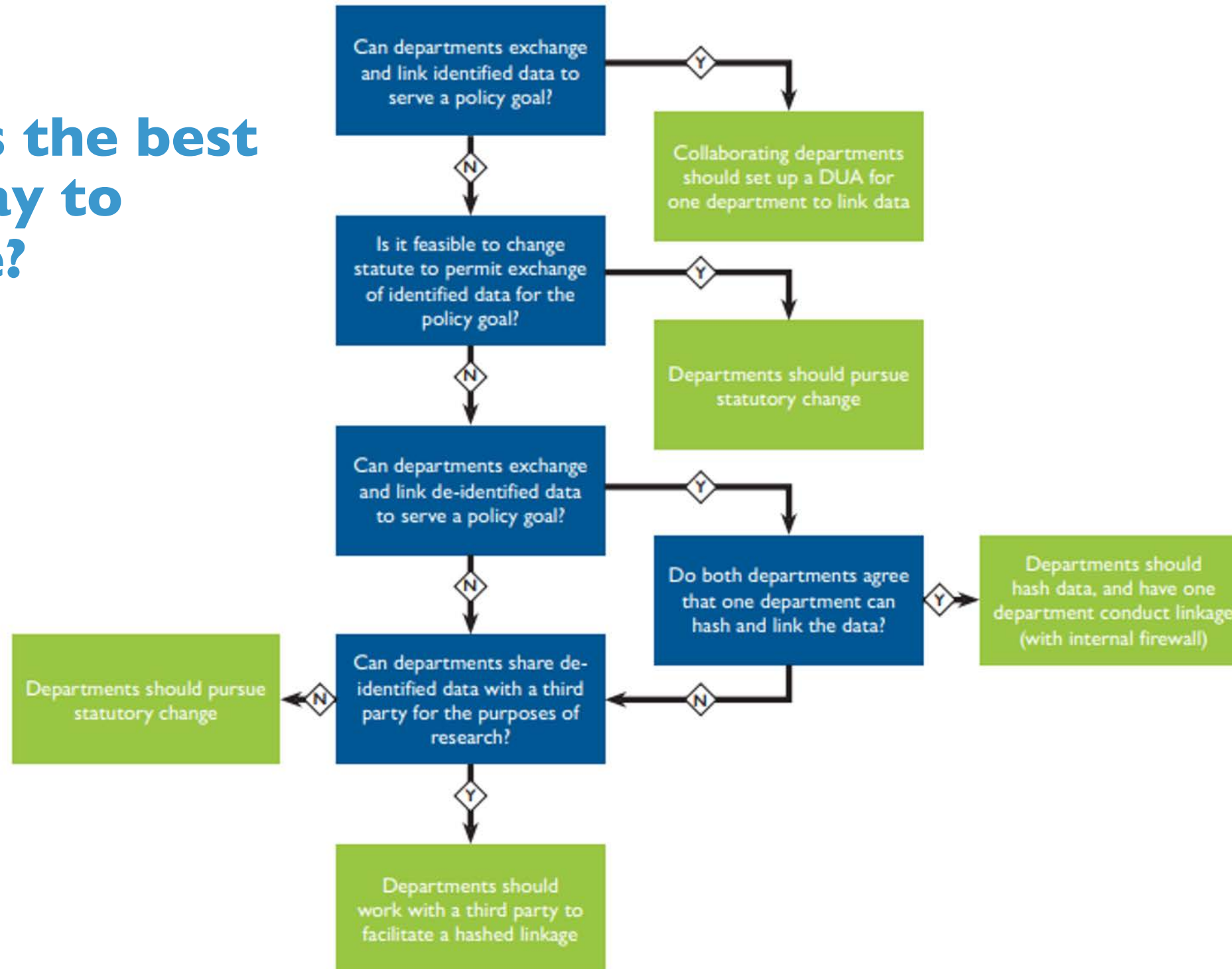
7

Ensure data
security

8

Technical teams
execute a
hashed linkage

What's the best pathway to linkage?



Hashed linkages...



- Take time
- Take resources
- Should only be used if there are no other viable pathways to linkage
- Can often provide a powerful proof of concept: helping departments make a case for why they should link data directly



Other linkages CPL is conducting

- 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



Questions?

evanbwhite@berkeley.edu

bit.ly/HashedLinkage