October 24, 2024

Government Data of the People, By the People, For the People

Navigating Citizen Privacy Concerns

Claire McKay Bowen, Ph.D
Senior Fellow and Data Governance and Privacy Practice Area Lead
✉ cbowen@urban.org 🌐 www.clairemckaybowen.com 🐦 @ClaireMKBowen

URBAN INSTITUTE · ELEVATE · THE · DEBATE

In Conversation with Matt Desmond:
How We Can End Poverty in America

#LiveAtUrban

The Senate has confirmed the 1st Latino to lead the U.S. census, Robert Santos

Updated November 4, 2021 · 5:56 PM ET

HANSI LO WANG

Robert Santos, president of the American Statistical Association, has been approved to lead the U.S. Census Bureau as its new Senate-confirmed director through 2026.

Tom Williams/CQ Roll Call via Getty Images

Features : : Spatial Equity Data Tool

Home   Documentation

Spatial Equity Data Tool

Measure Disparities in Your Data

Last updated March 19, 2024

Are libraries evenly located across the country? Do all residents in your state—regardless of race or income—have equitable access to polling places? Does your dataset accurately reflect your county's or city's population?

This tool can help government agencies, policymakers, and community advocates easily answer such questions by assessing demographic and spatial disparities in their data.

Choose a geography below to get started. Then check out our sample datasets or upload your own data.

National-level data   State-level data   County-level data   City-level data

Background
Promise of Tiered Data Access

Approach #1: restricting access

Common access restrictions

- Direct or restricted access

  o Federal Statistical Research Data Centers

  o NIH-supported data repositories

- User agreements

  o Data Use Agreements (DUA)

  o Non-Disclosure Agreements (NDA)

  o Memorandum of Understanding (MOU)

Why not apply for full access to the data?

Yes: In an idealized world, trusted researchers and policymakers could have direct access to the data.

Not everyone can:

- Eligibility requirements
  (e.g., must be a U.S. citizen).

- Lengthy clearance
  (e.g., months or years).

- Limited access to secure enclaves
  (e.g., hundreds of miles away).

Why not apply for full access to the data?

Yes: In an idealized world, trusted researchers and policymakers could have direct access to the data.

Not everyone can:

- Eligibility requirements (e.g., must be a U.S. citizen).

- ~~Lengthy clearance (e.g., months or years).~~

- ~~Limited access to secure enclaves (e.g., hundreds of miles away).~~
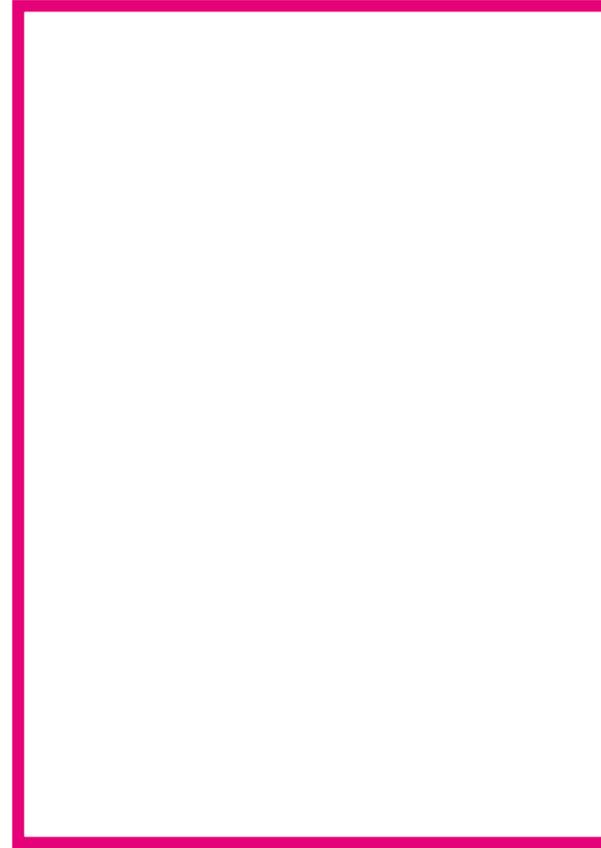

Secure Transfer, Restricted-Use Data Lake

Approach #2: querying statistics

How do we access confidential data?
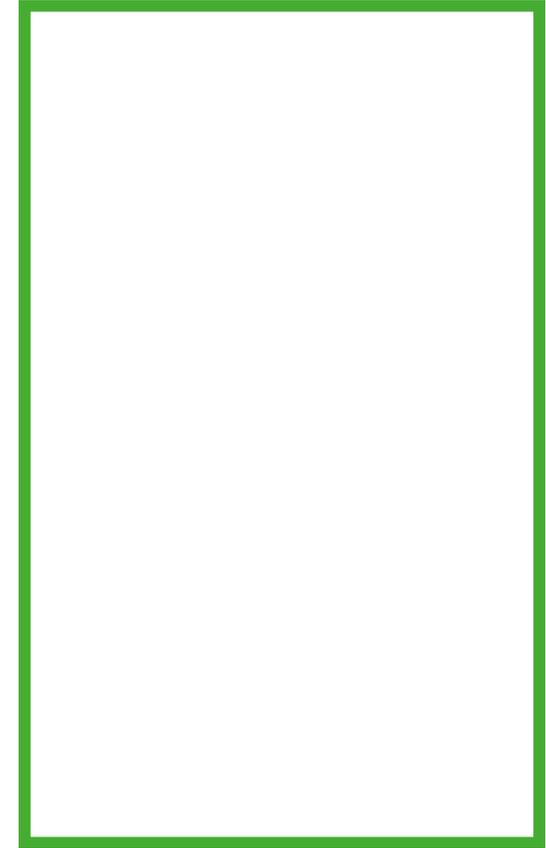
User Interface
Layer

Privacy
Layer

Data Access
Layer

URBAN INSTITUTE
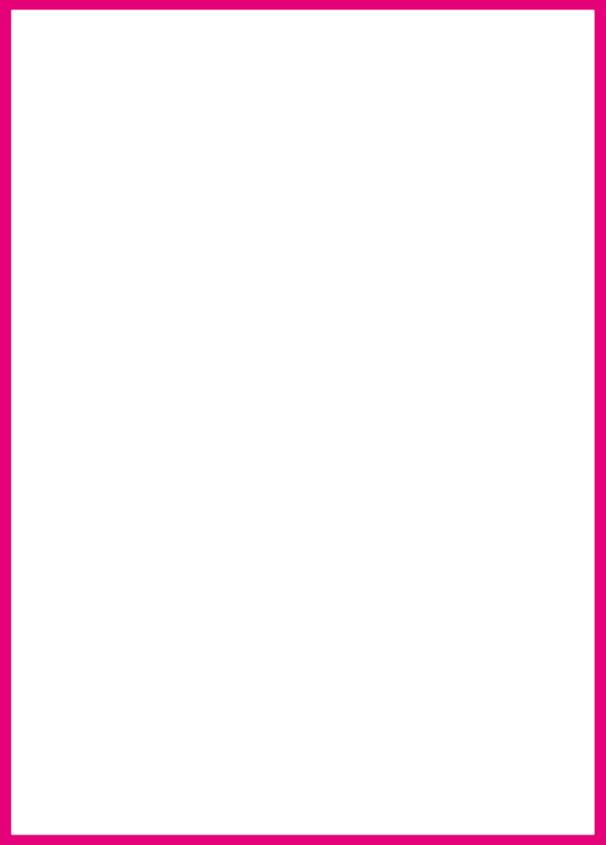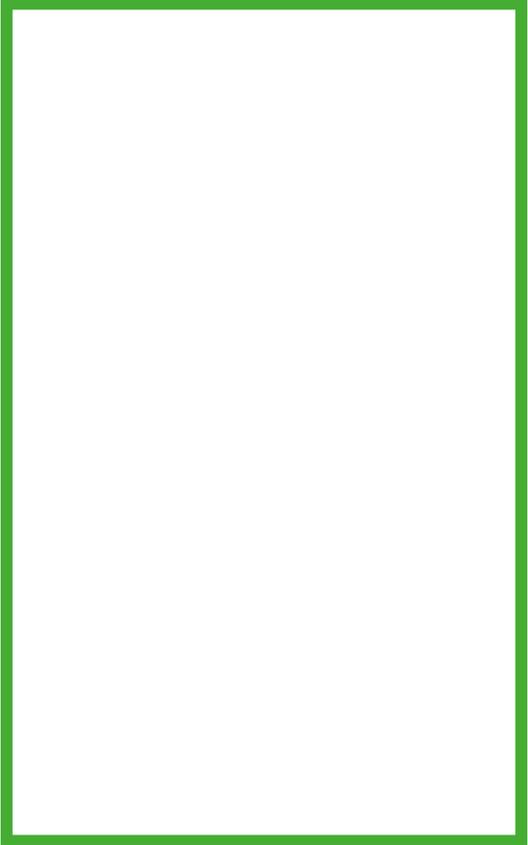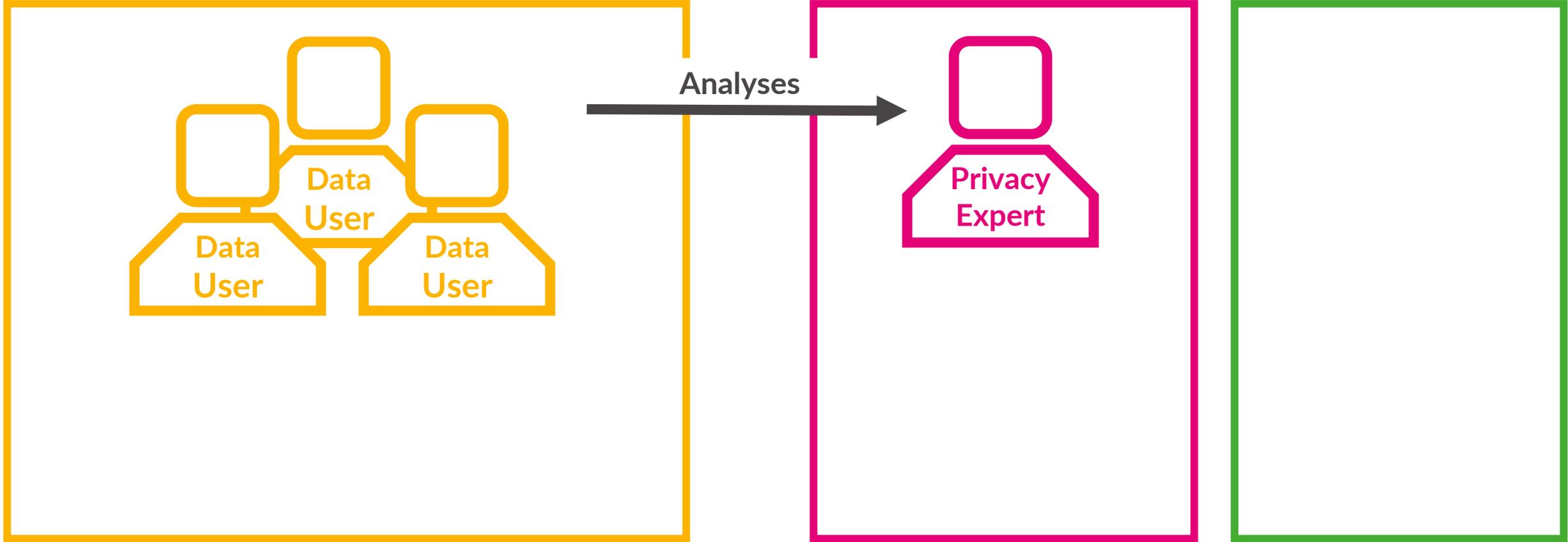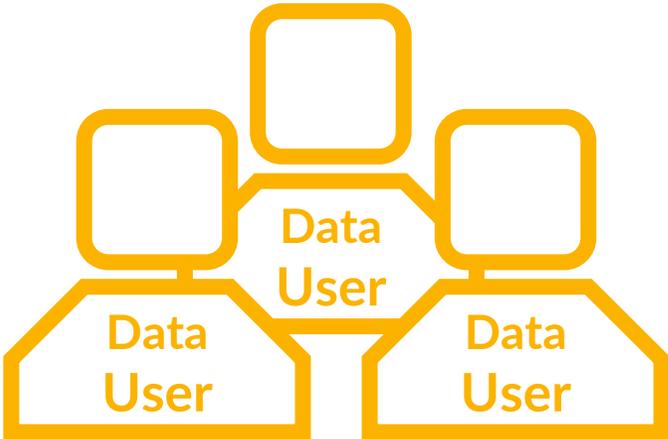
How do we access confidential data?

User Interface
Layer

Privacy
Layer

Data Access
Layer

Data User

Data User

Data User

URBAN · INSTITUTE

How do we access confidential data?

User Interface
Layer

Privacy
Layer

Data Access
Layer

Data
User

Data
User

Data
User

Analyses →

Privacy
Expert

· U R B A N · I N S T I T U T E ·

How do we access confidential data?

User Interface Layer → Privacy Layer → Data Access Layer

Data User, Data User, Data User → (Analyses) → Privacy Expert → (Analyses) → Data Curator, Conf. Data
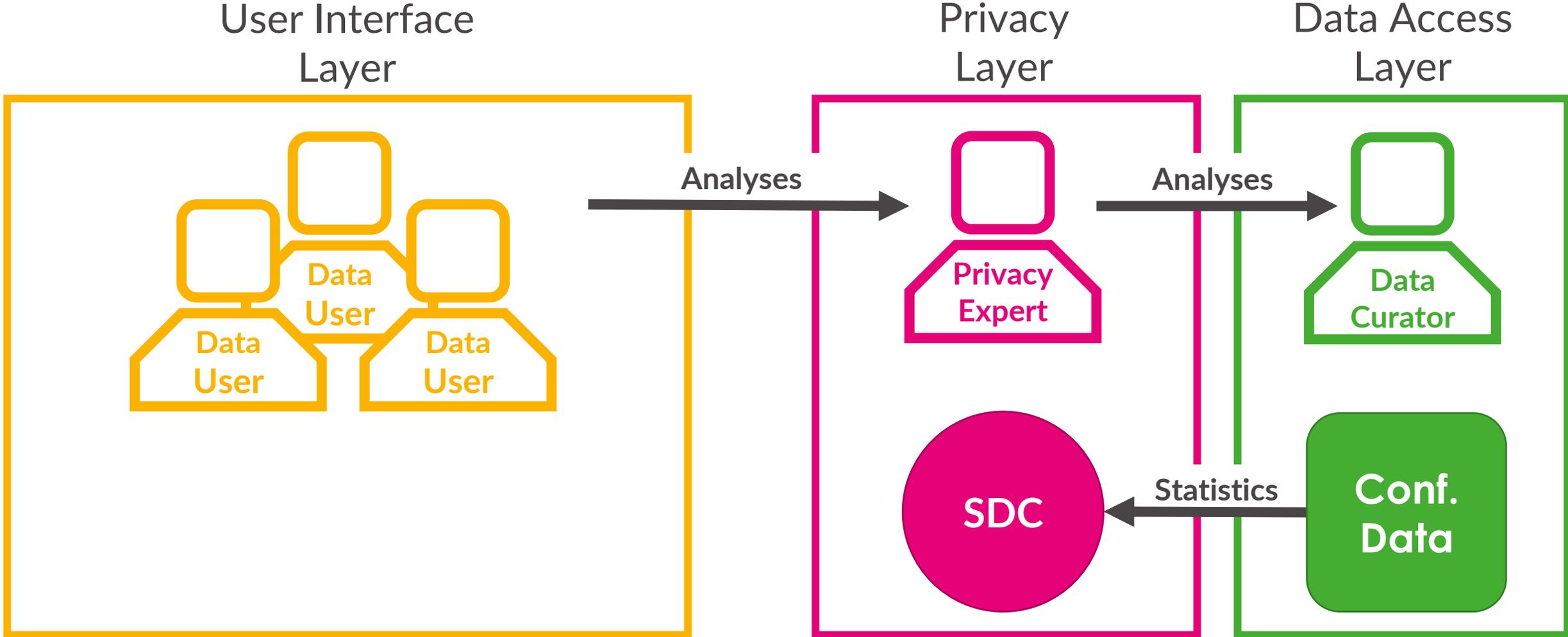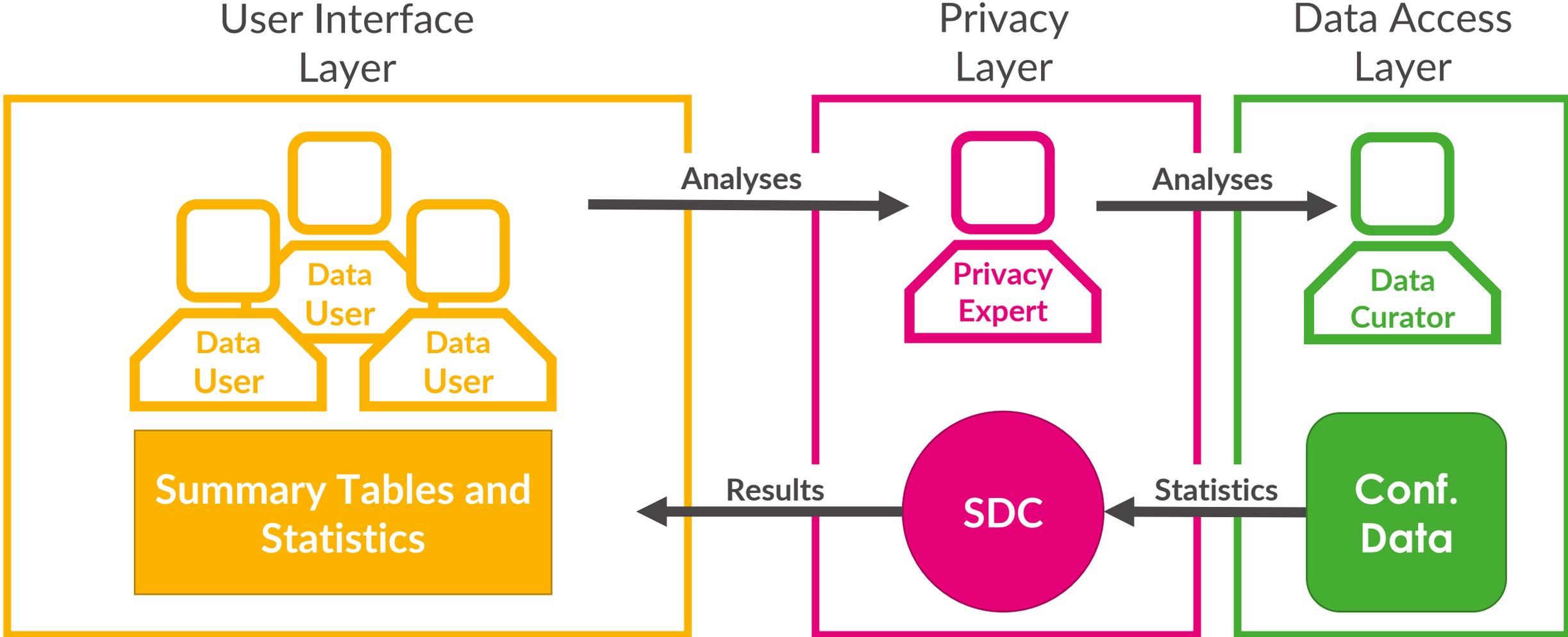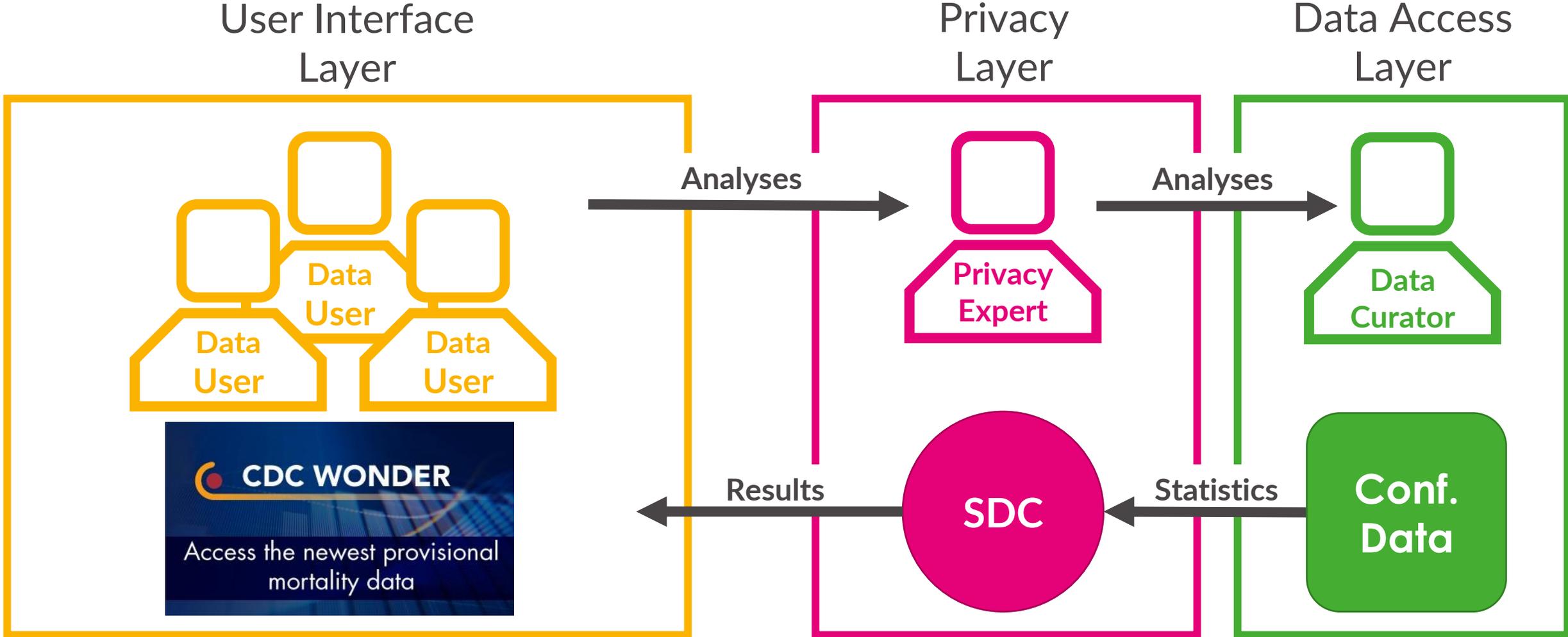
How do we access confidential data?

SDC / SDL = Statistical Disclosure Control or Limitation

How do we access confidential data?

SDC / SDL = Statistical Disclosure Control or Limitation

URBAN · INSTITUTE

How do we access confidential data?

User Interface Layer — Data User, Data User, Data User — CDC WONDER: Access the newest provisional mortality data

Analyses →

Privacy Layer — Privacy Expert

Analyses →

Data Access Layer — Data Curator

← **Results** — SDC ← **Statistics** — Conf. Data

SDC / SDL = Statistical Disclosure Control or Limitation

URBAN·INSTITUTE

User Interface Layer

Data User
Data User
Data User

Analysis 1
Analysis 2
Analysis 3
Analysis 4
Analysis 5
Analysis 6

Privacy Layer

Access Confidential Tax Data to Advance Your Research

+ NEW SCRIPT

DASHBOARD   ABOUT   HELP   ACCOUNT ▾

Submit Script   Review & Refine Initial Results   Release Results
1   2   3

Release Final Results

? Instructions

• Below is a list of all the cells in this analysis to which you have made refinements. Select the version of each cell you want to release.

• If you would like to make additional refinements before you make your selection, click "Refine Again" to return to step 2.

View More ⌄

Script Name › First Analysis   See All Analyses

ANALYSIS 1   PRIVACY VS. NOISE

PRIVACY BUDGET   PENDING COST TO RELEASE = 0
Review & Refine Budget   94.6 of 100   Public Release Budget   100 of 100
                                                                RELEASE

Apply privacy-preserving methodologies

Analysis 1.1

Analysis 1.2

noise

Analysis 1

Data Layer

Analyses →

Analyses →

← **Results**

← **Statistics**

Approach #3: publishing microdata

How do we access confidential data?

SDC / SDL = Statistical Disclosure Control or Limitation

What are ways to protect data?

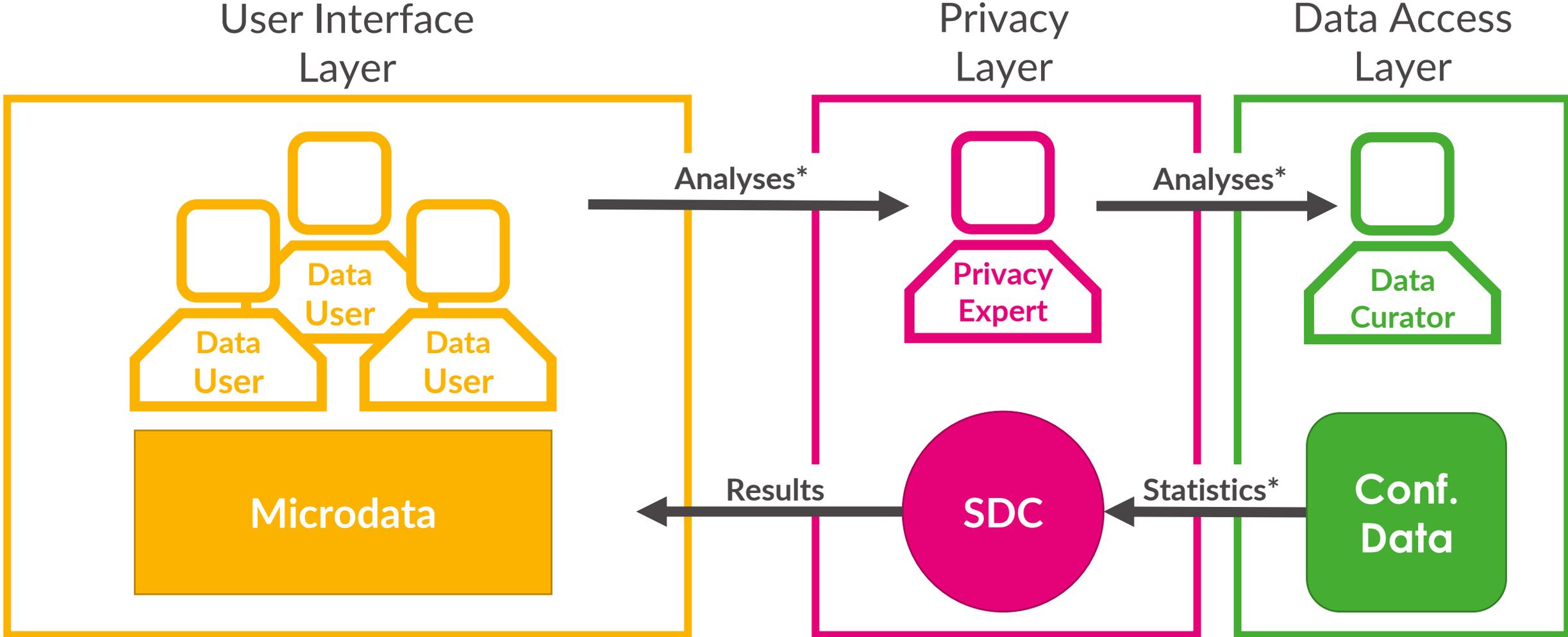What are ways to protect data? **Suppression**

What are ways to protect data? Generalization
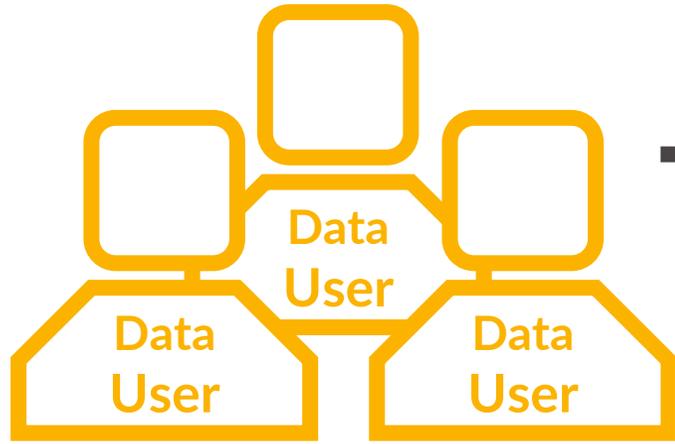
What are ways to protect data? Top Coding

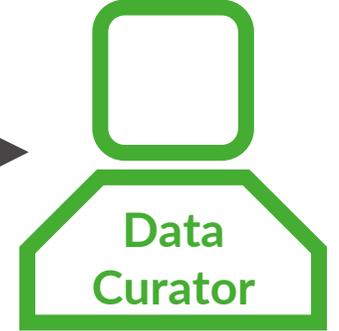What are ways to protect data? Synthetic Data

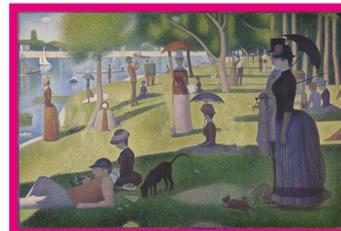# User Interface Layer

Privacy Layer

Data Layer

Analyses →

Privacy Expert

Analyses →

Data Curator

Apply privacy-preserving methodologies

← **Results**

← **Statistics**

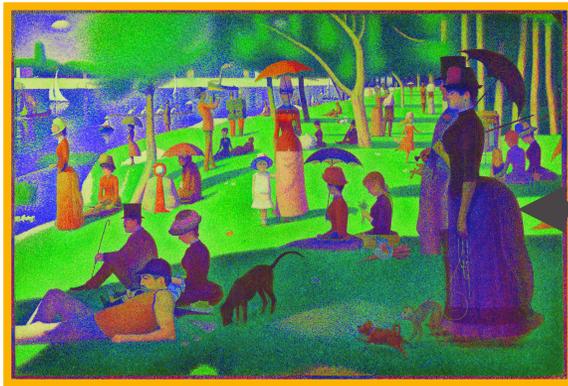**Context:**
park with a lake

Statistics:
people, color, grass, trees, water, sky
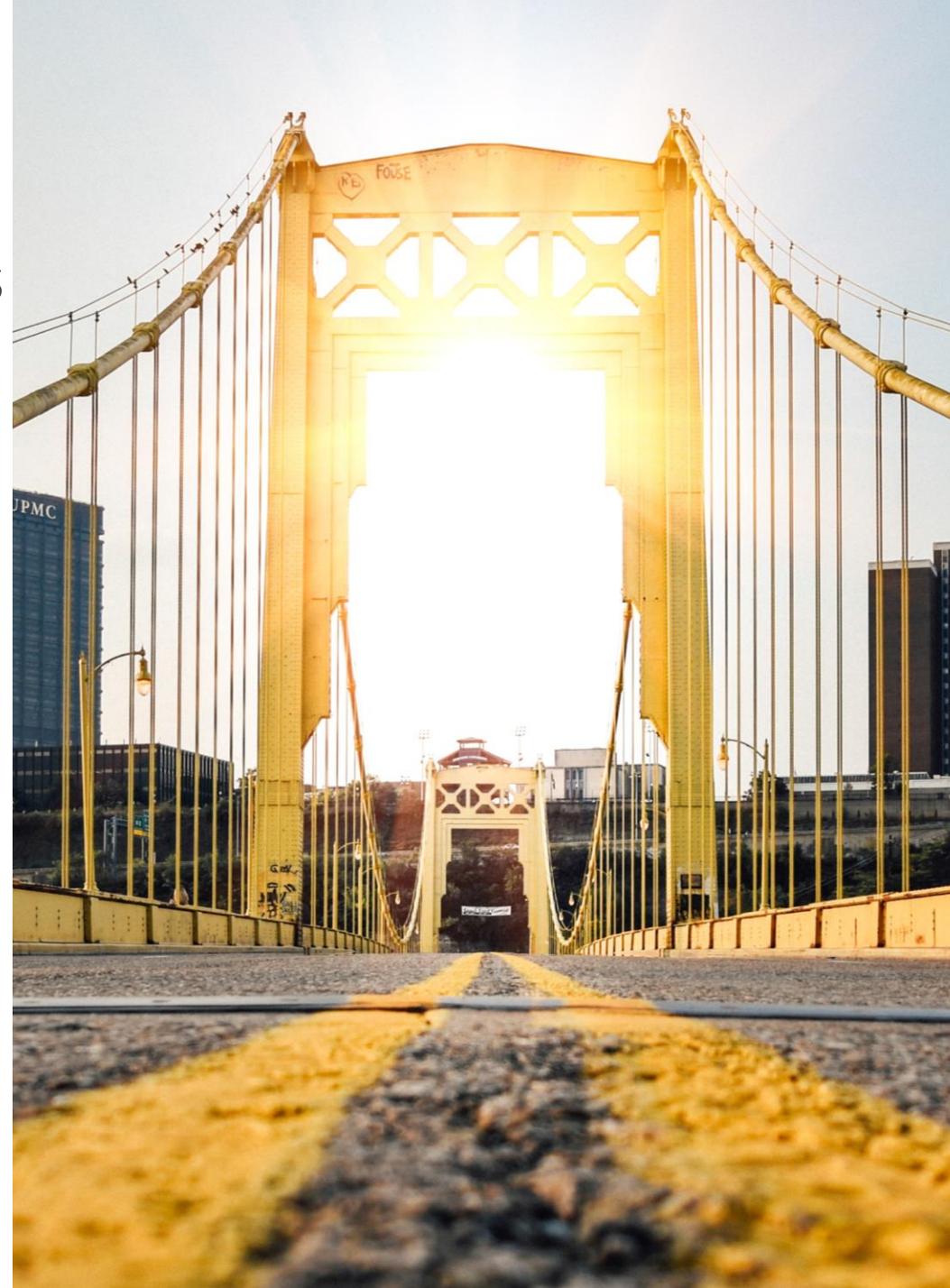
Urban Synthetic Data Applications

- Disaggregated synthetic human services data for Allegheny County

- Synthetic Nebraska state education and workforce data

- Synthetic courthouse data

- Synthetic Survey of Earned Doctorates

- Synthetic taxpayer data

- REAL Incubator Fund demonstration of opt-in disclosure protection

Contact Me

✉ cbowen@urban.org

🌎 www.clairemckaybowen.com

in /in/bowenclaire

🐦 @ClaireMKBowen