October 24, 2023

Data Privacy Considerations for Data Disaggregation

Striking a balance between data disaggregation and data privacy – technical considerations

Motivation

Why should anyone care?


ONE NATION, TRACKED
AN INVESTIGATION INTO THE SMARTPHONE TRACKING INDUSTRY FROM TIMES OPINION

HEALTH

'It's not a pretty picture': Why the lack of racial data around COVID vaccines is 'massive barrier' to better distribution

Nada Hassanein USA TODAY

Published 5:30 a.m. ET Feb. 1, 2021 | **Updated 2:10 p.m. ET Feb. 1, 2021**

Abigail Echo-Hawk, chief research officer with Seattle Indian Health Board and a member of the Pawnee Tribe, gets a shot of the Moderna COVID-19 vaccine on Dec. 21. A colleague used a black pen to inscribe "For the (Heart) love of Native People" over the injection spot. *Karen Ducey, Getty Images*

JOURNAL REPORTS: TECHNOLOGY

No Place to Hide: Colleges Track Students, Everywhere

Schools use tech to follow students online, on the quad and in the football stadium

By Douglas Belkin Follow

March 5, 2020 10:02 pm ET

BRIEFING ROOM

Executive Order On Advancing Racial Equity and Support for Underserved Communities Through the Federal Government

JANUARY 20, 2021 • PRESIDENTIAL ACTIONS

"...advanc[e] equity for all, including people of color and others who have been historically underserved, marginalized, and adversely affected by persistent poverty and inequality."

FEBRUARY 16, 2023

Executive Order on Further Advancing Racial Equity and Support for Underserved Communities Through The Federal Government

 ▸ BRIEFING ROOM ▸ PRESIDENTIAL ACTIONS

To iterate, expand, further coordinate the critical work of embedding equity in government-wide processes.

Why does this matter (now)?

- Increase of data and computing power, which increases the risk of re-identifying individuals in survey and administrative data.
- Ability to measure privacy risks and protect against them is evolving rapidly.
- Survey response rates declining, more interest in using administrative data for research.

Gaps Exist Between Regulations and Practical Realities

- Regulations exist e.g.,
  - HIPAA
  - Title 13
  - GDPR; CCPA

- Cannot be clearly translated to technical methods
  - Do not define de-identification or minimal privacy risks
  - More focused on collection than dissemination
  - Requires interpretation/discretion


THE PROMISE OF EVIDENCE-BASED POLICYMAKING
Report of the Commission on Evidence-Based Policymaking
Transparency
Humility
Data
Privacy
NATIONAL ACADEMIES Sciences Engineering Medicine
Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good
Consensus Study Report
ACDEB
Advisory Committee on Data for Evidence Building
Advisory Committee on Data for Evidence Building: Year 2 Report
October 14, 2022

Background

Promise of Tiered Data Access

Approach #1: restricting access

Commons access restrictions

- Data enclaves
  - Federal Statistical Research Data Centers (link)
  - NIH-supported data repositories (link)
- User agreements
  - Data Use Agreements (DUA)
  - Non-Disclosure Agreements (NDA)
  - Memorandum of Understanding (MOU)

Why not apply for full access to the data?

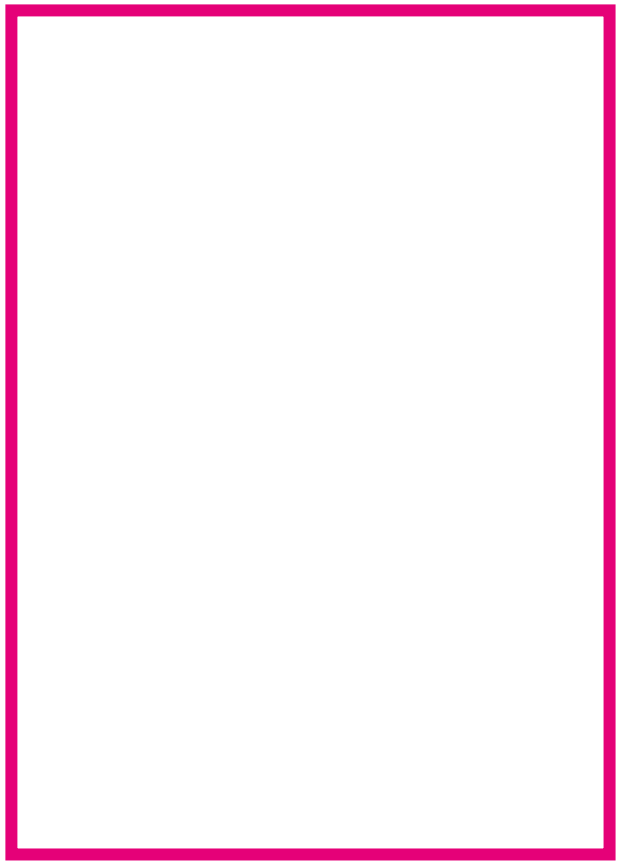
- **Yes:** In an idealized world, trusted researchers and policymakers could have direct access to the data.
- **Not everyone can:** There are several hurdles such as:
  - Eligibility requirements (e.g., must be a U.S. citizen).
  - Lengthy clearance process to gain direct access (e.g., months or years).
  - Limited access to Federal Research Data Centers (e.g., hundreds of miles away).

Approach #2: querying statistics

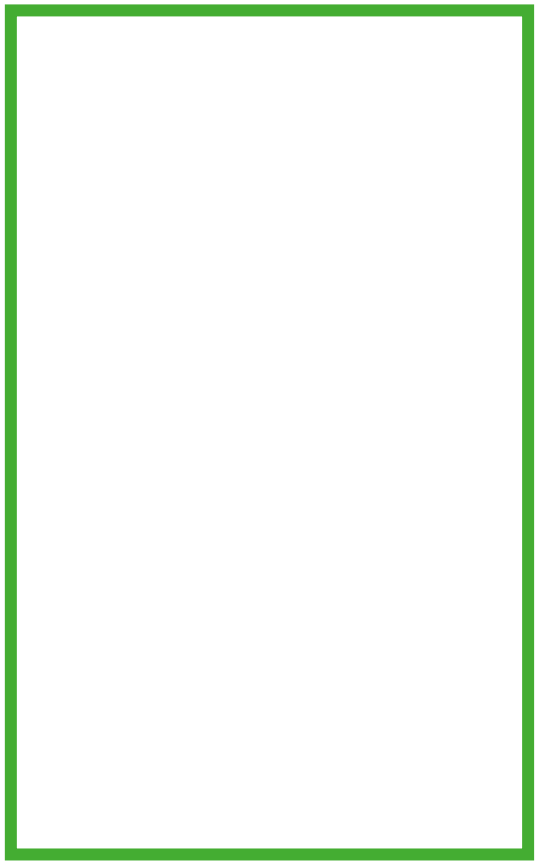
How do we access confidential data?

User Interface Layer

Privacy Layer

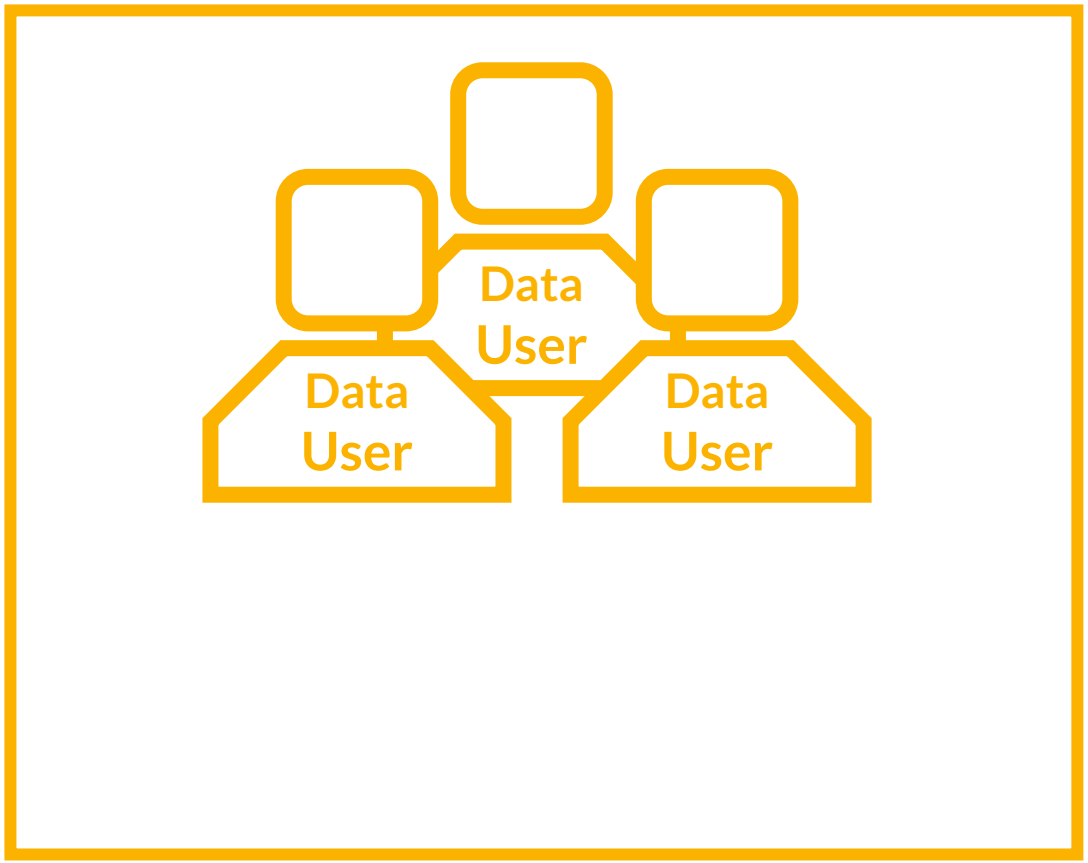


Data Access Layer



How do we access confidential data?

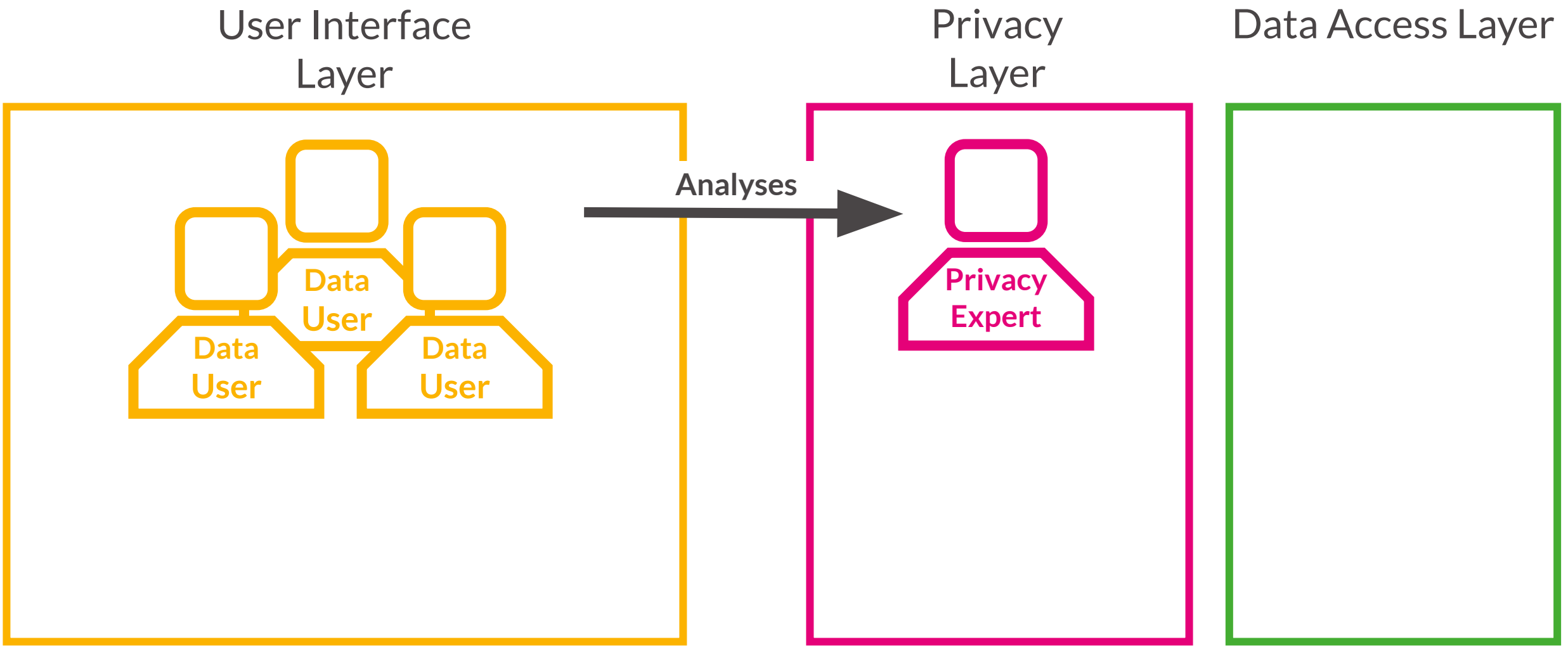
User Interface Layer



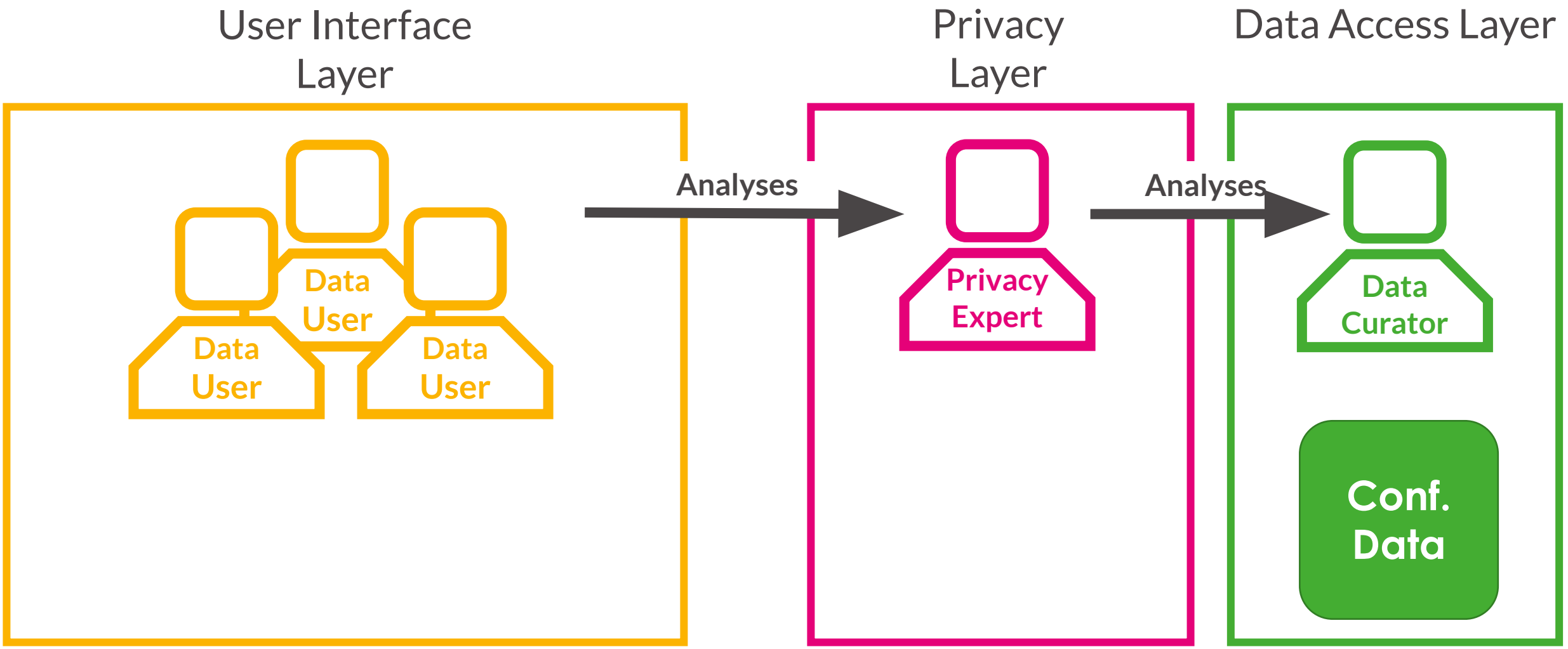
Privacy Layer

Data Access Layer

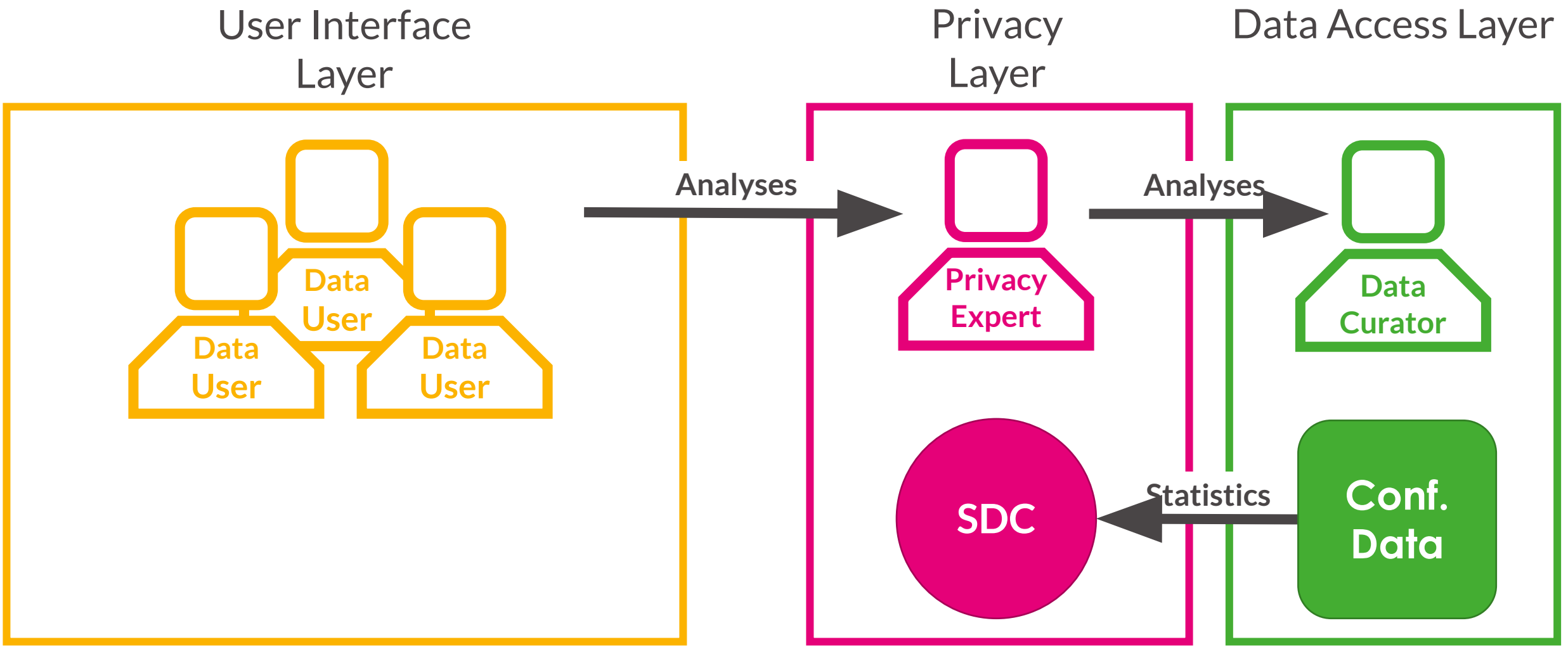
How do we access confidential data?



How do we access confidential data?

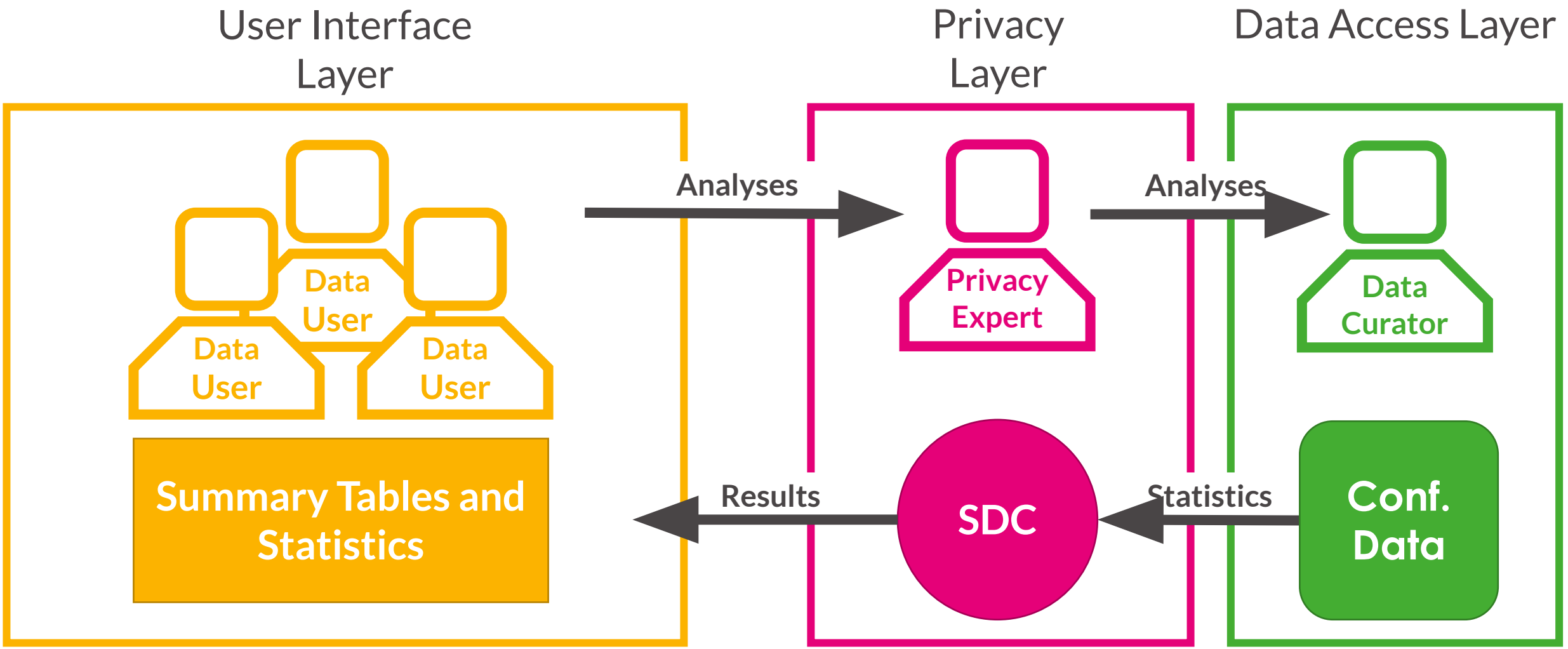


How do we access confidential data?



SDC / SDL = Statistical Disclosure Control or Limitation

How do we access confidential data?



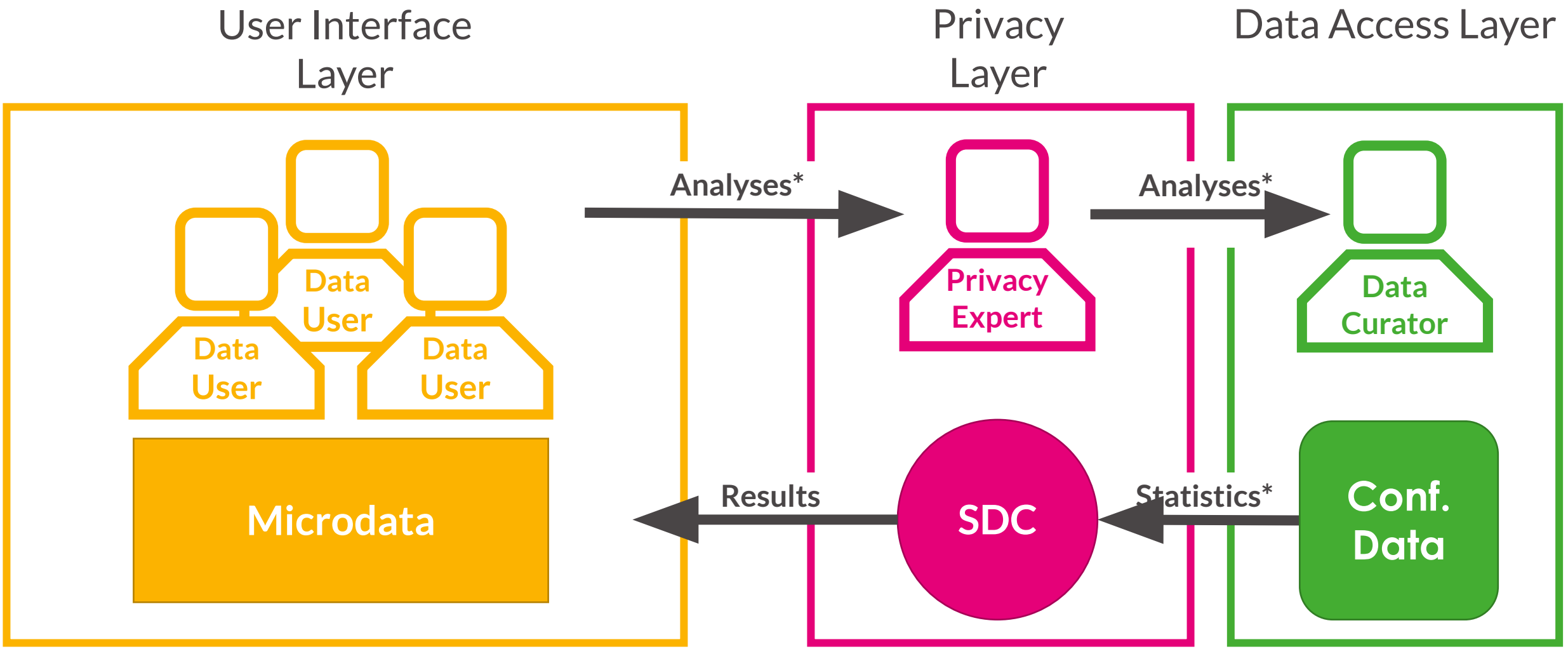
SDC / SDL = Statistical Disclosure Control or Limitation

Frameworks for querying statistics

- Predefined systems
  (e.g., table builders)
- Validation servers
  (e.g., submit R code to run a regression model directly on the confidential data)
- Verification servers
  (e.g., submit an analysis run on public data and receive model "verification“)

Approach #3: publishing microdata

How do we access confidential data?



SDC / SDL = Statistical Disclosure Control or Limitation

Tradeoffs of publishing microdata

- Benefits
  - Easier to control the privacy loss
  - More familiar for data users
  - Less infrastructure needs
- Drawbacks
  - Harder to capture the use cases
  - Less targeted process
  - Allows for downstream privacy loss

Contact Me

cbowen@urban.org

www.clairemckaybowen.com

/in/bowenclaire

@ClaireMKBowen