2024 FCSM Research & Policy Conference

Communicating Quality While Building a Privacy-Preserving Validation Server

Safe Data Technologies: Safely Expanding Access to Administrative Tax Data

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Erika Tyagi
Lead Data Engineer, Urban Institute

Project Funding & Collaborators

ALFRED P. SLOAN
FOUNDATION

Project Goal

> *The Safe Data Technologies project aims to safely expand access to confidential data that advances evidence-based policy making by creating new ways for researchers to use administrative data while protecting privacy.*

Traditional Access for Administrative Tax Data

I. **Basic Access:** Summary tables on the Statistics of Income Division website.

II. **Full Access:** Researchers who obtain clearance and therefore have access to the unaltered, confidential data, but will be still be limited on what information can be released.

Tiered Access for Administrative Tax Data

Our goal is to enable more researchers to safely access confidential tax data.

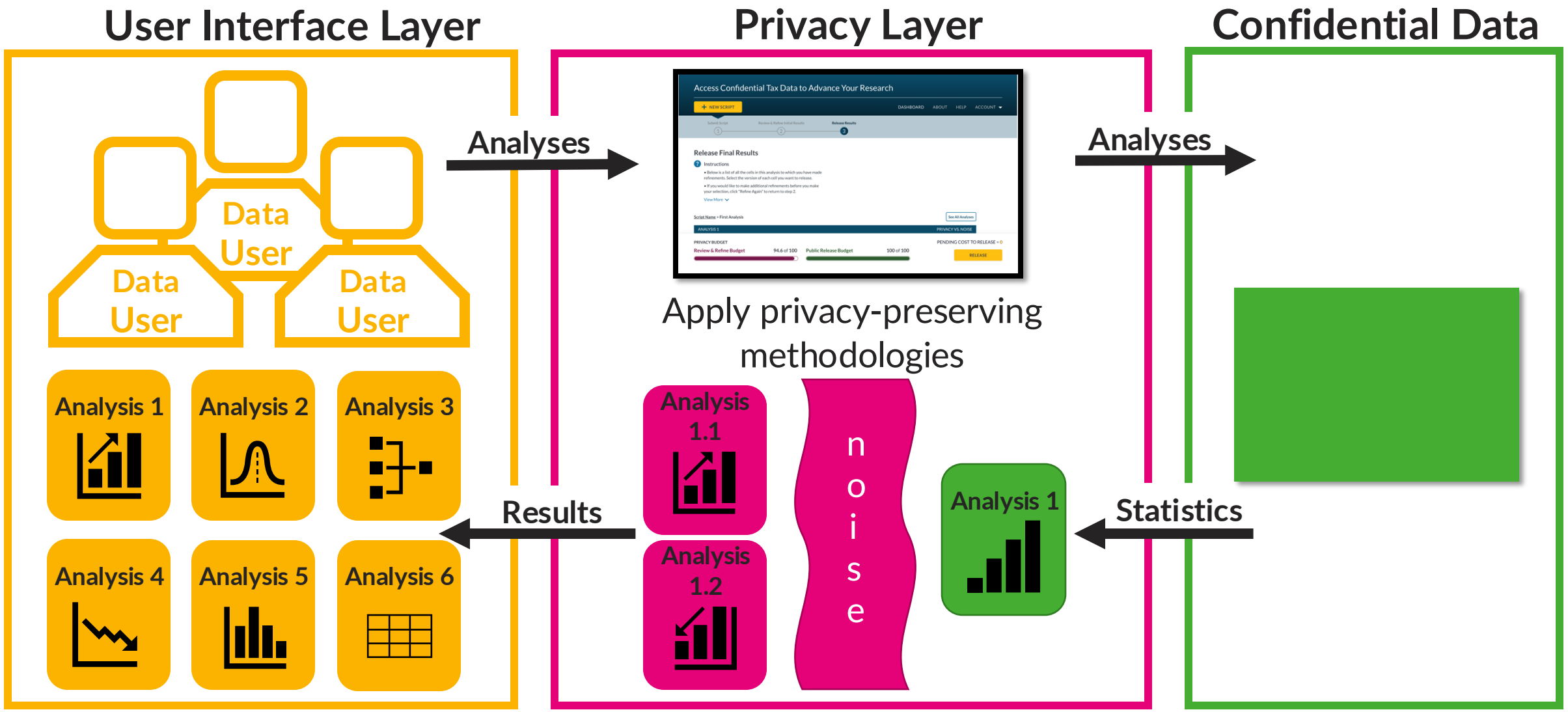
I. **Basic Access:** Summary tables on the Statistics of Income Division website.

II. **Public Use File (PUF):** Researchers will have access to the synthetic PUF.

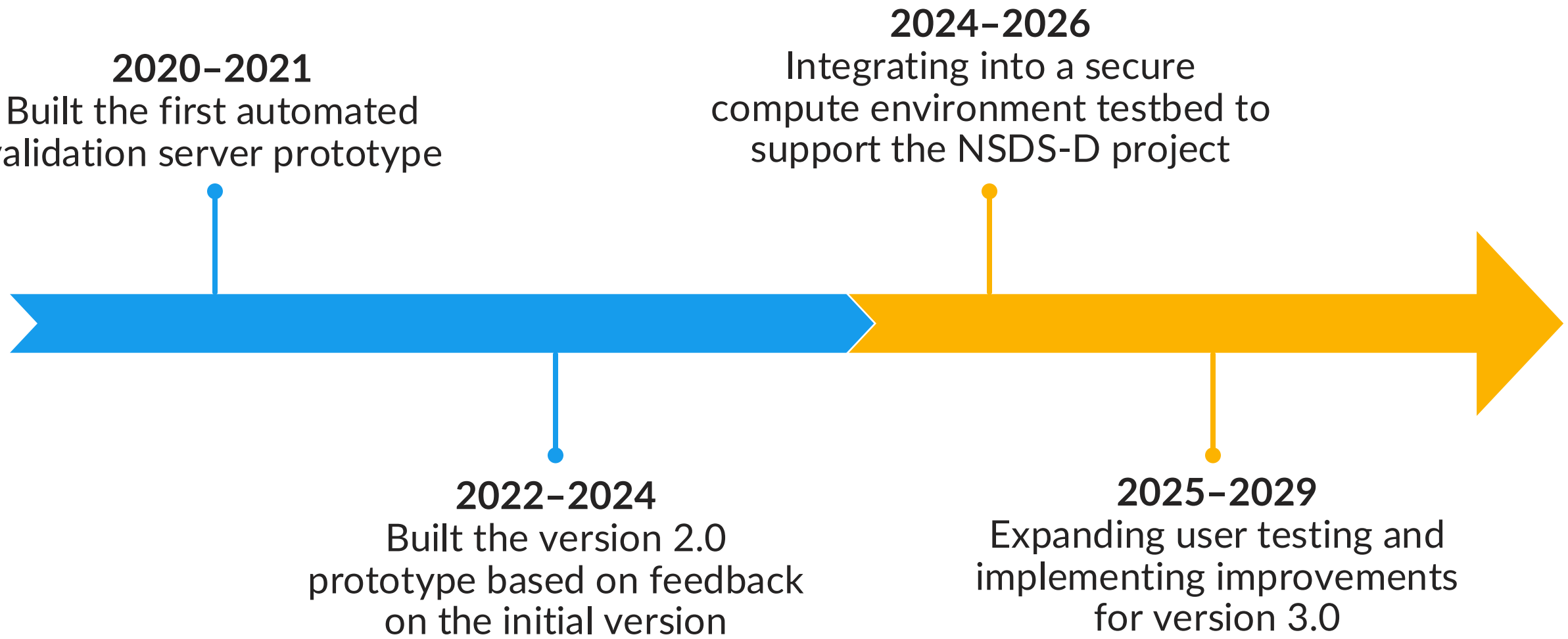
III. **Validation Server Access:** Researchers are trusted to access the validation server, where they submit statistical analyses that they have tested and debugged on the synthetic PUF. Researchers at this tier will have to undergo an application process.

IV. **Full Access:** Researchers who obtain clearance and therefore have access to the unaltered, confidential data, but will be still be limited on what information can be released.

Validation Server: Access Mechanism



Validation Server: Prototype Development History



Traditional Stakeholders in Data Privacy

- Data privacy experts
- Data users or practitioners
- Data stewards and agency staff
- Data participants or subjects

Stakeholders While Developing a Validation Server

- Data privacy experts
- Data users or practitioners
- Data stewards and agency staff
- Data participants or subjects
- Software engineers
- Designers and user experience experts
- Security experts
- IT and infrastructure experts

Examples of Tensions in Communicating Quality

1. How can we develop a system that resembles researcher workflows?

Examples of Tensions in Communicating Quality

1. How can we develop a system that resembles researcher workflows?
2. Which privacy algorithm(s) should we implement?

Examples of Tensions in Communicating Quality

1. How can we develop a system that resembles researcher workflows?
2. Which privacy algorithm(s) should we implement?
3. What do terms like “security” and “risk” mean?

Lessons Learned

1. Reiterate first principles: What are we doing and why?

Lessons Learned

1. Reiterate first principles: What are we doing and why?
2. Define terminology.

Lessons Learned

1. Reiterate first principles: What are we doing and why?
2. Define terminology.
3. Prioritize bridging the gap between theory and practice.

Lessons Learned

1. Reiterate first principles: What are we doing and why?
2. Define terminology.
3. Prioritize bridging the gap between theory and practice.

Successfully developing tools to create new tiers of access for administrative data will require **empathy** and **thoughtful communication** across a growing set of partners in the federal data ecosystem.

Contact Us & Learn More

safedatatech@urban.org

Validation Server Version 2.0 White Paper

Safe Data Technologies Project Landing Page