URBAN INSTITUTE

A Privacy-Preserving Validation Server Version 3.0

Technical White Paper for an Updated Automated Validation Server Prototype

Erika Tyagi, Silke Taylor, Graham MacDonald, Deena Tamaroff, Claire McKay Bowen, and Joshua Snoke

May 2026

This technical white paper provides an overview of improvements made to the Urban Institute's privacy-preserving validation server 3.0 prototype. Our goal in designing the validation server is to provide a practical solution for government agencies looking to improve and automate their statistical disclosure processes, enabling potential users to analyze confidential data without direct access to it. By continuing to develop this tool, we hope to significantly increase access to valuable data and insights used to craft better public policy.

In the first section, we summarize the purpose of the project, key concepts, and our previous work developing earlier prototypes. In the second section, we describe our process for gathering feedback from potential users, data stewards, and privacy experts, including formal user testing. We then detail the improvements implemented in version 3.0 and enhancements to the system's infrastructure in the fourth section. We conclude with a discussion of potential future directions for the project.

Purpose of the Project

The Safe Data Technologies project is a multi-year effort led by the Urban Institute, in collaboration with partners at the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS), to design and test new approaches for making sensitive administrative tax data more accessible for research while preserving privacy and utility. Our work responds to the growing recognition that traditional disclosure control methods are increasingly inadequate in the face of modern re-identification risks and often slower in releasing of valuable data (Burman et al. 2024). By developing and piloting innovative tools that protect privacy while maintaining analytic usefulness, we aim to help data stewards, such as SOI, expand access to invaluable information in ways that are both safe and practical for informing evidence-based policymaking (Bowen et al. 2024).

The collaboration's primary strategy is to create **synthetic data**—data that imitate a confidential dataset while limiting information about individual records—and accompany the release of that synthetic data with a validation server. While synthetic data generally produces accurate results for certain analyses, data users may wish to validate their results against the confidential data itself if the synthetic file are not designed for certain analyses. The validation server allows them to do so, in an automated manner, without ever viewing the confidential data itself.

The Safe Data Technologies project has generated synthetic data, improved synthetic data generation tools, and a first-of-its-kind automated privacy-preserving validation server to fulfill this goal. At a high level, after a data user submits their analysis to the validation server, the server adds noise to returned statistics and associated standard errors. To account and manage the disclosure risk for all the analyses across multiple projects and users, the validation server uses a formal privacy-inspired mechanism. In combination with synthetic data products, the validation server provides agencies with a new tier of data access that can be more efficient, reliable, and robust for many use cases.

In our previous white paper on version 2.0 of the validation server prototype, we described the design and functionality of the system in detail (Taylor et al. 2021; Tyagi et al. 2024). In this brief, we build on that work, summarizing improvements to the system's infrastructure, user interfaces, and privacy protections, as well as enhancements informed by feedback from privacy experts and user testing.

Process for Conducting User Studies and Soliciting Feedback

In September 2024, we released a technical white paper describing version 2.0 of the validation server prototype, which improved significantly upon its predecessor developed in 2021 (Tyagi et al. 2024). Our goal with the version 2.0 prototype was to deploy a fully operational system that could enable broader user testing with potential data users and data stewards and to identify the improvements needed for agency adoption.

Following the release, we immediately engaged a wide range of stakeholders to gather feedback for the next iteration. We presented the updated prototype at several conferences, including the 2024 Federal Committee on Statistical Methodology Conference and 2024 Privacy and Public Policy Conference. In February 2025, we convened our advisory board to review the system's current state, challenges, and potential enhancements. Our advisory board focused on the validation server consists of 24 experts across a range of fields, including statistics, computer science, economics, and demography. They provided extensive feedback on improvements to support adoption and build trust in the technology, which we then summarized to update our development roadmap and priorities for the version 3.0 prototype. In particular, our advisory board members emphasized the following:

- Clearly demonstrating that the validation server and broader process within which it sits (including release of a synthetic public use file and publishing outputs from the system) would be faster and easier for data users and government agencies
- Simplifying the privacy budget concept and implementation to improve accessibility and usability
- Conducting extensive and ongoing red-teaming and disclosure risk assessments to build trust in the system
- Providing educational materials and training to support widespread adoption

In addition to this feedback, we conducted 18 user testing sessions between February 2025 and April 2025 to further assess the usability of the prototype with data users, data stewards, and privacy experts. Led by the Urban Institute's Lead User Experience Specialist, participants were asked to navigate the workflow and provide feedback on areas where visual elements or steps were ambiguous, unintuitive, or could be improved.

Across these sessions, the following key takeaways emerged:

- Users wanted more explanatory text throughout the interface to provide context on unfamiliar concepts (e.g., privacy budgets, epsilon values). They also struggled interpreting the trade-off between noise and privacy loss.
- Users found the separate "review and refinement" and "release" privacy budgets confusing, and privacy experts raised concerns about the implications of maintaining both.
- Users wanted the ability to add and update descriptions for submissions and refinements after submission to help distinguish multiple versions.
- Users wanted to validate the syntax of their scripts without drawing from their privacy budgets.
- Users requested project-based access management, enabling multiple data users to collaborate more easily through the system.

In the following section, we describe how these findings shaped the improvements implemented in the version 3.0 validation server prototype.

Key Improvements of the Version 3.0 Prototype

The version 3.0 prototype introduces several key improvements: a newly developed administrator interface, privacy and statistical validity enhancements recommended by experts, and a suite of usability updates based on feedback from the user studies described in the previous section.

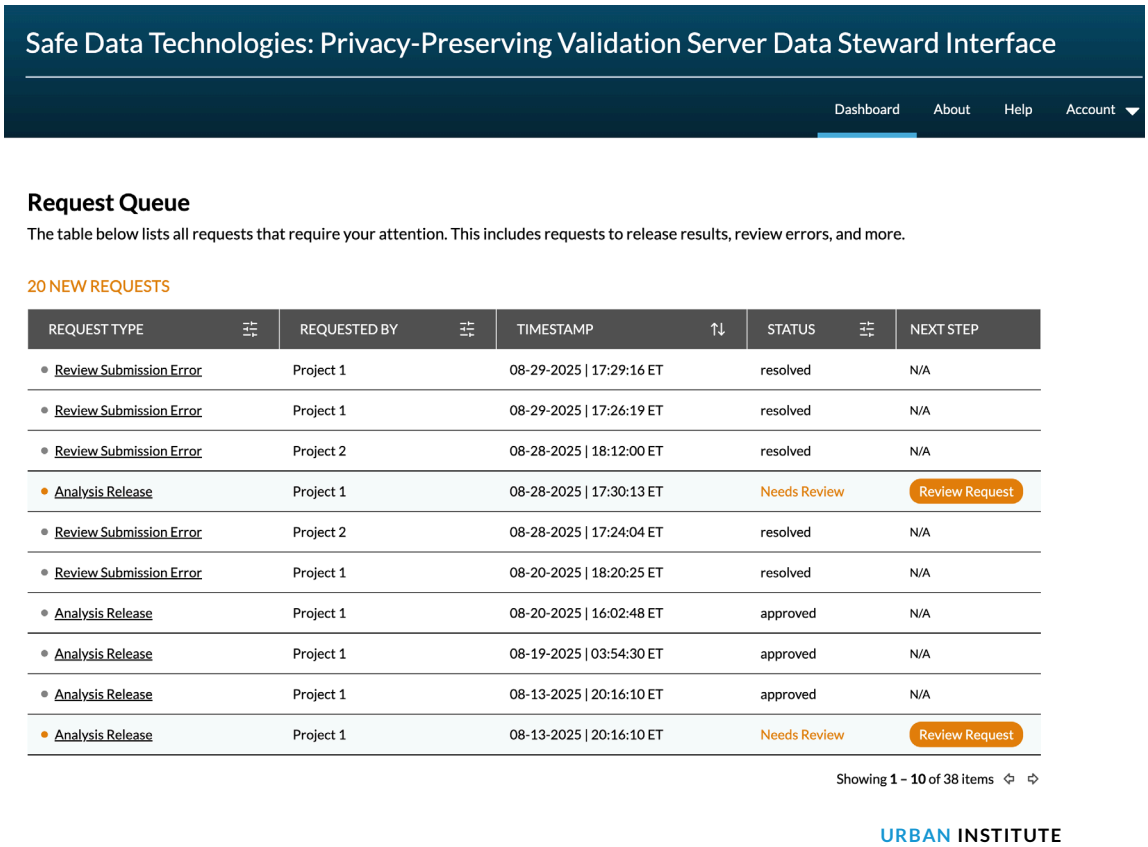
Administrator Interface

To make the management of user submissions simpler and more seamless, the version 3.0 system includes a dedicated interface for administrators (or data stewards) to manage and monitor activity within the validation server. Through this interface, administrators can review users and projects with access to the system, monitor and adjust privacy budgets, review requests to release results or disclose error messages, and audit activity to understand disclosure risks associated with their data assets.

FIGURE 1

Validation Server Version 3.0 Administrator Interface Landing Page

Landing page showing the administrator's request queue



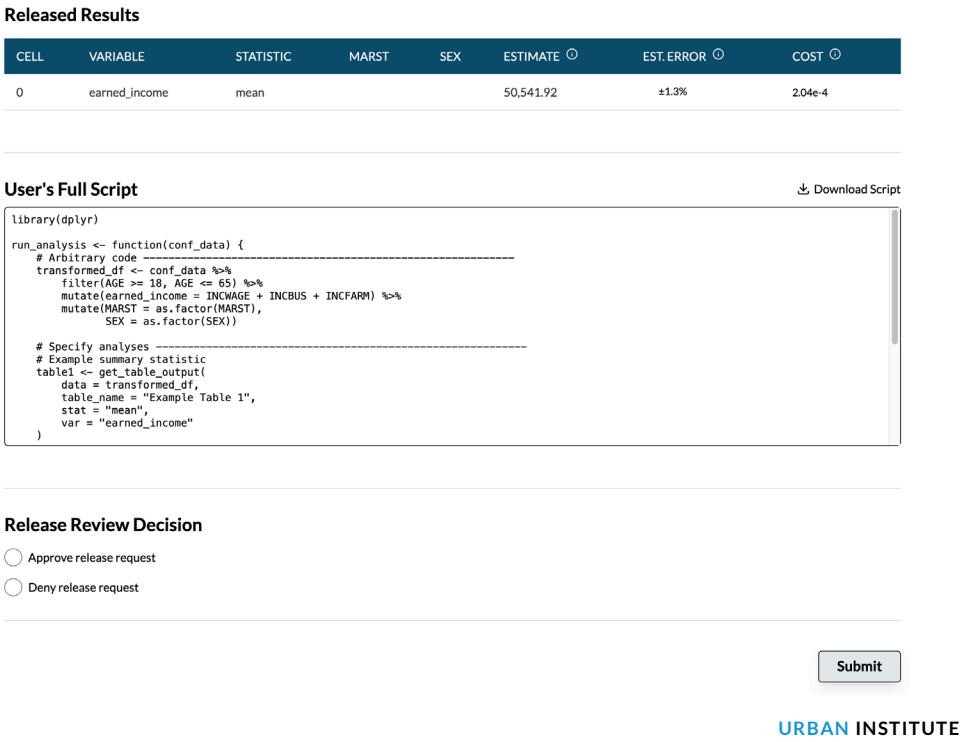
Source: Authors' screenshot of the interface.
Notes: Upon logging into the administrator interface, data stewards see a landing page with their request queue. This view summarizes actions such as analysis release requests, submission error reviews, privacy budget updates, and more.

As with the user interface, we adopted a user-centered design approach. We began by gathering input on the key features administrators would want from the system. We then developed and tested wireframes with potential data stewards, including staff at federal and state statistical agencies, to assess usability and intuitiveness. Following several rounds of iteration, we then implemented the final designs and deployed the application.

FIGURE 2

Validation Server Version 3.0 Administrator Interface Release Review Page

Administrator interface for reviewing requests to release results from the secure environment



Source: Authors' screenshot of the interface.
Notes: Users can submit requests to release results from the secure environment for publication or broader dissemination. This interface allows administrators to review these requests and decide whether to approve or deny them. Decisions are then communicated back to users.

Privacy and Statistical Validity Enhancements

Improved Statistical Validity of Regression Output

As discussed in the version 2.0 white paper, the validation server uses a generalized form of a local sensitivity approach, inspired by the Maximum Observed Sensitivity (MOS) algorithm proposed by Chetty and Friedman (2019), to add noise to tabular and regression outputs, where the noise is based on the data unlike formally private methods. In the earlier prototype, we computed sensitivities independently for each statistic from an analysis and then added the noise accordingly. While this provided flexibility and scalability by allowing the same privacy algorithm to support a broad range of

outputs, it resulted in inconsistent relationships among regression statistics and their associated standard errors, limiting their validity for statistical inference. In other words, in version 2.0, adding noise to the estimate and the associated standard error independently could lead to improper 95 percent confidence interval, a misleadingly large t-statistic, and a spurious finding of statistical significance—an artifact of the privacy algorithm rather than a true result from the underlying data.

In version 3.0, we revised the method to add noise only to a subset of base statistics for regressions (coefficients, residual sum of squares, total sum of squares, and number of observations), still following the MOS-inspired approach. All other regression output statistics are derived from these base statistics after noise is applied, preserving their internal consistency and ensuring validity for statistical inference. This means that the t-statistic in our example above would now reflect the true level of uncertainty in the data.

Improved Approach to Handling Errors in User-Submitted Scripts

In the version 2.0 white paper, we discussed the challenge of balancing usability and privacy when handling errors in user-submitted scripts. Analysts will inevitably submit R scripts with errors (e.g., syntax issues, misspelled variable names), even if they develop their code using a synthetic or public-use file. Because certain errors can reveal sensitive information (such as whether specific observations exist), some privacy experts recommend withholding error messages in interactive query systems and stochastically returning errors to the user even when their code did not contain an actual error (Barrientos et al. 2018). However, potential users overwhelmingly indicated that this approach would make the system frustrating and, in many cases, unusable.

In version 3.0, we adopted a hybrid approach. Uploaded scripts are first run against a synthetic dataset that mirrors the structure (variable names and types) of the confidential data. If execution fails on the synthetic data, the error message and traceback are returned directly to the user. If the script runs successfully on synthetic data but fails on the confidential dataset (e.g., due to empty cell errors after filtering), the error is routed to the administrator via the new administrator interface. The administrator can then decide whether to share the message or provide alternate guidance based on the potential disclosure risk.

This human-in-the-loop process is not a perfect solution, but we believe it strikes a workable balance for vetted users. We will continue refining this approach through deployments with trusted testers, using insights on error frequency and type to inform future enhancements.

Improved Protection against Edge Cases

Even if we do not publish the local sensitivity value, recent work by Protivash and colleagues (2024) demonstrates vulnerabilities of local sensitivity approaches. In summary, their paper suggest that threshold function attacks can exploit cases where local sensitivity is zero, resulting in no noise being added to a statistic. This situation can occur, for example, when a user filters the data to a subset where a variable has no variation. To mitigate this risk, version 3.0 of the validation server enforces a non-zero sensitivity for every statistic by specifying a minimum sensitivity of 1. This ensures that all outputs include some noise, even when the observed data would otherwise imply zero sensitivity.

In addition, privacy experts have highlighted risks that arise when users allocate most of their privacy budget on a single statistic, reducing noise to negligible levels. To address this, we introduced explicit upper and lower bounds on privacy loss (epsilon) values across all requested statistics.

These measures help strengthen the system against edge-case vulnerabilities while maintaining usability for legitimate research use. We recognize that long-term protection requires ongoing monitoring of the privacy literature, regular red teaming, and governance structures that hold vetted users accountable for appropriate use. We also hope that training and support for data stewards can allow them to better understand and monitor for suspicious activity, enabling earlier detection of potential misuse and stronger safeguards in practice.

User Experience Enhancements

Supports Multiple Datasets and Project Workflows

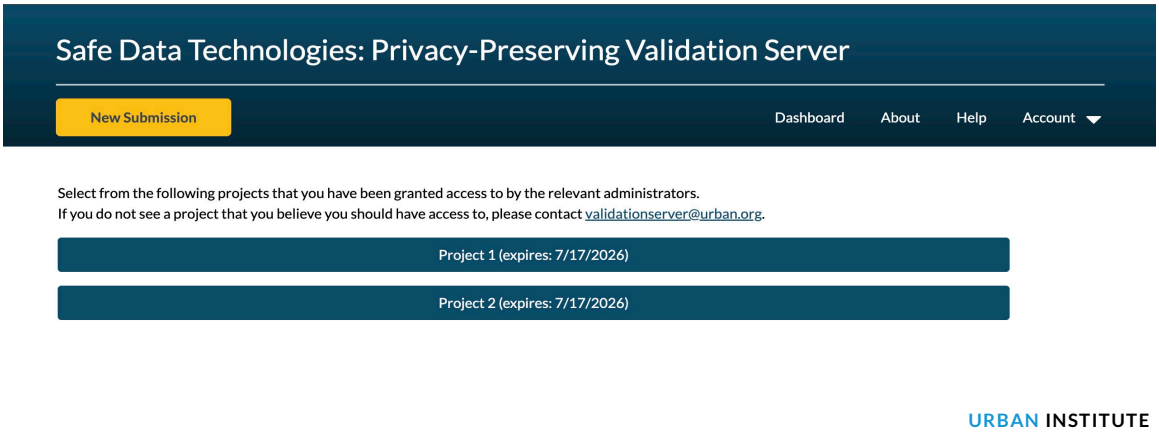
When engaging with potential users and adopters of a validation server across conferences and user studies, two requests frequently came up to support deployment in research settings: support for multiple datasets and functionality for multiple users to collaborate on a shared project through the interface. Both features are now implemented in the version 3.0 prototype.

When a user logs into the validation server, they are prompted to select from a list of projects they have access to, each associated with its own dataset. Once a project is selected, the interface reflects that choice: submitted scripts run against the relevant dataset, each project maintains an independent privacy budget, and only submissions tied to the selected project are displayed. Multiple users can be assigned to a project, sharing the same privacy budget and viewing each other's submissions.

FIGURE 3

Validation Server Version 3.0 Project Selection Interface

User interface for selecting from available projects



URBAN INSTITUTE

Source: Authors' screenshot of the interface.
Notes: When a user logs into the validation server, they are prompted to select from a list of projects they have access to. Each project is associated with its own dataset and, optionally, a set of collaborations.

Consolidates Privacy Budgets

In user studies, both potential users and data stewards found the previous "review and refinement" and "release" privacy budgets confusing. They reported that tracking multiple budgets created more burden than benefit. Data stewards at government agencies, particularly those without deep privacy expertise, were unsure how to set and manage two separate budgets or interpret their relative disclosure risks. Privacy experts also expressed concern about the unclear implications of the intermediate "review and refinement" budget and how to explain it to data stewards. In version 3.0, each project now has a single privacy budget.

Version 2.0 had these separate budgets because, in previous test sessions, users expressed a desire to submit some intermediate analyses to learn more about the data before submitting a final request for release and publication. Agency data stewards can still accommodate this use case if they would like by providing sufficient privacy budget, but they will not have to manage or contend with the difference between two different budgets.

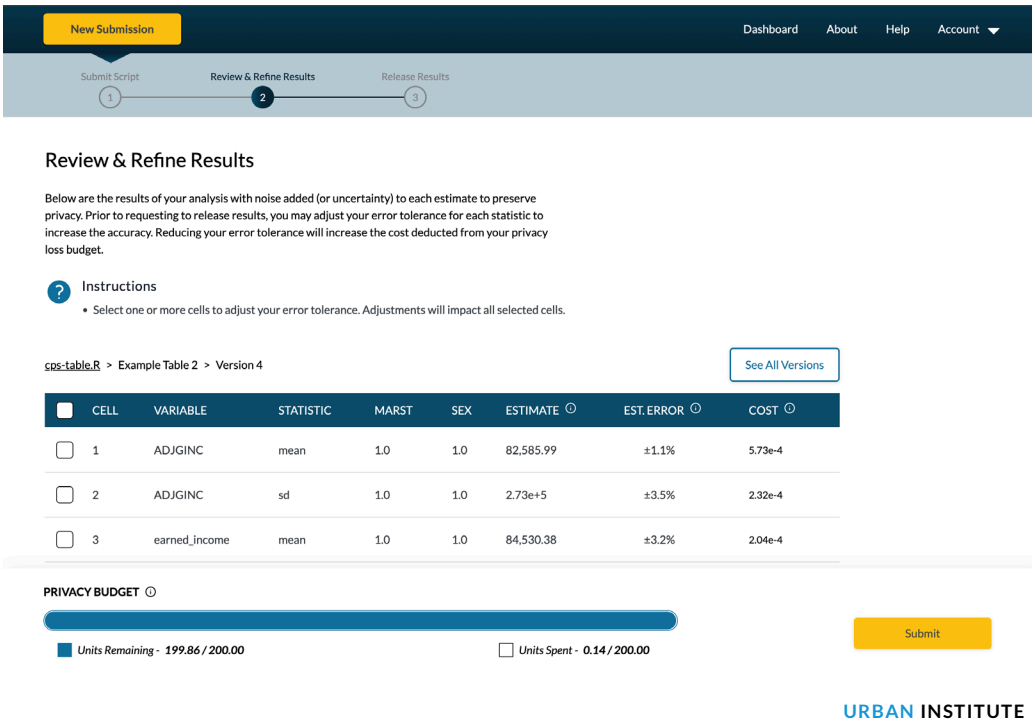
Allows Users to Specify Error Tolerances

Potential users in user testing also found it challenging to interpret and set epsilon values for the privacy loss budget. While they broadly understood that there was trade-off between added noise and privacy cost, they struggled to translate this into appropriate values for their analyses or context. However, they were more comfortable expressing tolerable levels of statistical accuracy (for example, allowing up to 10 percent error for a regression coefficient). In response, the interface now lets users specify an approximate error tolerance directly. The system uses this to calculate the privacy budget cost, which is displayed immediately, enabling users to request refinements in familiar terms while understanding the privacy implications before submitting.

FIGURE 4

Validation Server Version 3.0 “Review and Refine Results” Interface

User interface to review and refine results from previously submitted analyses



URBAN INSTITUTE

Source: Authors’ screenshot of the interface.
Notes: The user can review and refine results from a successful submission. They can specify an approximate error tolerance, which the system then uses to calculate the privacy budget cost (displayed as the pending cost).

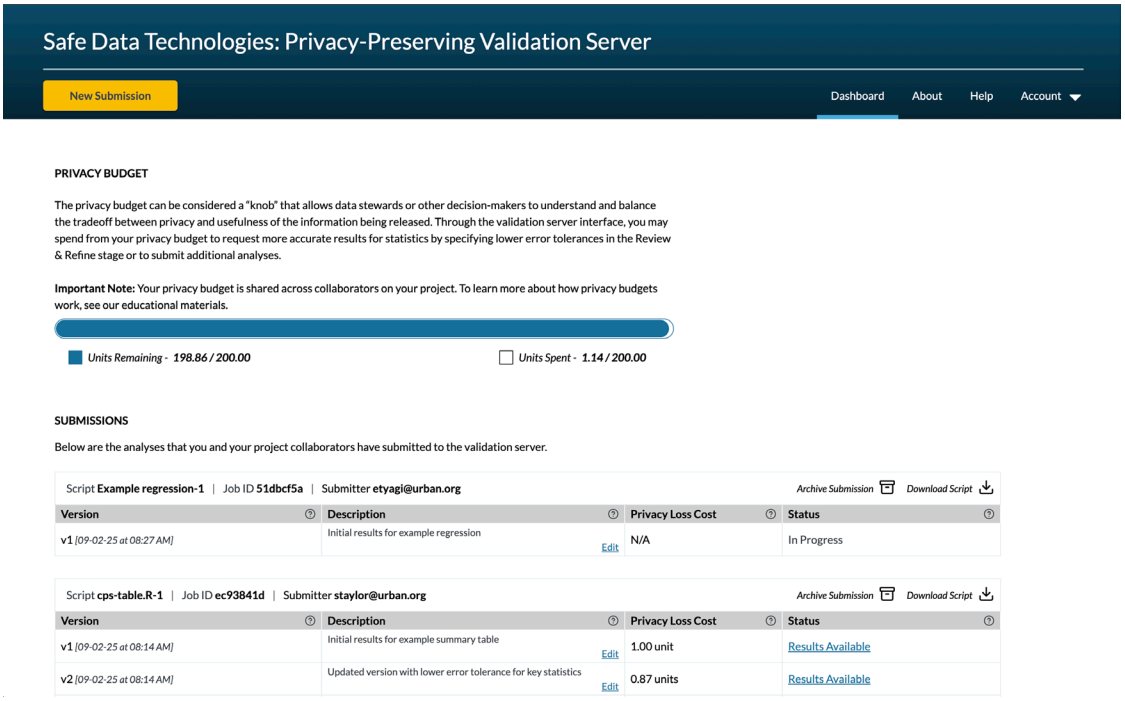
Adds Additional Display Features and Improves Explanatory Text

User testing also revealed that researchers wanted more ways to organize their analyses after submission. Currently, users assign names to analyses within their R scripts before submitting them and can also name each job (a group of analyses within a single R script) through the interface. However, several researchers noted that they would also like to retroactively add detailed descriptions to specific versions or refinements to help differentiate them. Users also suggested the ability to archive submissions to avoid cluttering their dashboard, especially when sharing the interface with multiple collaborators on a project. In the latest version of the tool, both features have been implemented through a redesigned landing page and submission dashboard.

FIGURE 5

Validation Server Version 3.0 Dashboard Interface

Landing page and dashboard with the user's privacy budget and previous submissions



URBAN INSTITUTE

Source: Authors' screenshot of the interface.
Notes: After selecting a project, the user can access a dashboard showing their remaining privacy budget and details about their previous submissions associated with that project. When multiple users are assigned to a project, they share a privacy budget and can view each other's submissions. This example shows the dashboard for a returning user with additional explanatory text and display features introduced in version 3.0.

Users also overwhelmingly indicated that additional explanatory text throughout the interface would help provide context and introduce concepts that may be unfamiliar, such as privacy budgets or epsilon values. In response, we added more detailed explanatory text and tooltips throughout the interface to clearly define terminology and remind users how these concepts are implemented in the system. This also aligns with the advisory board's recommendation for creating more materials to support training and adoption of the validation server.

Security and Scalability Enhancements

Like its predecessor, version 3.0 of the validation server is deployed on Amazon Web Services (AWS) using services that meet the highest level of federal security and compliance requirements. The latest release is hosted within an isolated AWS GovCloud environment to support potential adoption by government agencies. The application now runs in private subnets with no direct internet access, and all

updates and management traffic are routed through Virtual Private Cloud endpoints to further reduce the risk of security vulnerabilities.

Version 3.0 also adds single sign-on integration with multi-factor authentication required for all logins, including researchers accessing the researcher interface and data stewards accessing the administrator interface, managed through Amazon Cognito. We strengthened access controls and implemented security best practices across the development environment: backend logins now use machine-to-machine authentication; secrets are stored and automatically rotated using AWS Secrets Manager; and MFA is required to deploy updates to the production server through the continuous integration/continuous deployment (CI/CD) pipeline. Each environment (i.e., staging and production) has its own limited-access roles to more strictly separate permissions.

In addition to security enhancements, we improved the system's reliability and scalability to ensure readiness for production use. For example, we integrated load balancing to handle heavier traffic and now define and deploy all AWS resources (including networking components) using AWS CloudFormation infrastructure-as-code templates, ensuring that changes are tracked, versioned, and consistently reproducible across environments.

Future Directions

While any technology system has room for improvement, the authors believe that the validation server in its current iteration is ready for test deployment and adoption in federal, state, and local government agencies. In the coming one to two years, the team plans to focus its efforts on working with government agencies and initiatives to release synthetic data and deploy the validation server for testing with trusted external researchers. Throughout this process, Urban plans to continue to harden and certify the security of the validation server against strict federal government cybersecurity standards.

In addition, we hope to help agencies identify key privacy and disclosure officials who can be the owners of the synthetic data and validation server processes and systems, and support these officials to better understand the novel, automated nature of these systems and the tradeoffs inherent in measuring risk and disclosure within them.

And, of course, the team has a long list of user and privacy improvements and enhancements that we'd like to make to the server over time, as user understanding and demand for these tools increases and the state of the art in privacy enhancing technologies evolves. Some of those improvements include, but are not limited to, the following:

- Providing an automated report for the released statistics that includes example language on the validity of the outputs that can be used for external publications like policy reports, helping users understand and better communicate how the results were generated, what the noisy estimates mean, and how valid inference can be made

- Computing sensitivities directly where closed-form solutions exist (rather than through the MOS-inspired approach) to reduce computational intensity and return results more quickly for certain types of analyses
- Tracking privacy loss at a per-record level, based on whether each record is included in multiple queries, to more accurately capture true privacy loss across filtered subsets of data
- Expanding the range of supported analyses and programming languages, for example, by enabling more complex regression types (after ensuring that our approach to adding noise is suitable for these analyses) and allowing code submissions in Stata or Python in addition to R
- Offering more flexibility in implementing various privacy algorithms, for example by applying differentially private or other formally private methods for analyses well-suited to them (or when preferred by data stewards) and local sensitivity-based approaches for others

We plan to continue to make improvements to the system, interface, and underlying privacy algorithms over time.

References

Barrientos, Andrés F., Alexander Bolton, Tom Balmat, Jerome P. Reiter, John M. de Figueiredo, Ashwin Machanavajjhala, Yan Chen, Charley Kneifel, and Mark Delong. 2018. "Providing Access to Confidential Research Data through Synthesis and Verification: An Application to Data on Employees of the U.S. Federal Government." *The Annals of Applied Statistics* 12 (2): 1124–156. https://doi.org/10.1214/18-AOAS1194.

Bowen, Claire McKay, Joshua Snoke, Aaron R. Williams, and Andrés F. Barrientos. 2024. "The Case for Researching Applied Privacy Enhancing Technologies." Working Paper 32909. Cambridge, MA: National Bureau of Economic Research.

Burman, Leonard, Barry Johnson, Victoria L. Bryant, Graham MacDonald, and Robert McClelland. 2024. "Protecting Privacy and Expanding Access in a Modern Administrative Tax Data System." *National Tax Journal* 77 (4). https://doi.org/10.1086/732683.

Chetty, Raj, and John N. Friedman. 2019. "A Practical Method to Reduce Privacy Loss when Disclosing Statistics Based on Small Samples." Working Paper 25626. Cambridge, MA: National Bureau of Economic Research.

Panavas, Liudas, Joshua Snoke, Erika Tyagi, Claire McKay Bowen, and Aaron R. Williams. 2024. "But Can You Use It? Design Recommendations for Differentially Private Interactive Systems." *Preprint, arXiv*.

Protivash, Prottay, John Durrell, Daniel Kifer, Zeyu Ding, and Danfeng Zhang. 2024. "Reconstruction Attacks on Aggressive Relaxations of Differential Privacy." *Journal of Privacy and Confidentiality* 14 (3). https://doi.org/10.29012/jpc.871.

Taylor, Silke, Graham MacDonald, Kyle Ueyama, and Claire McKay Bowen. 2021. "A Privacy-Preserving Validation Server Prototype." Washington, DC: Urban Institute.

Tyagi, Erika, Silke Taylor, Graham MacDonald, Deena Tamaroff, Josh Miller, Aaron R. Williams, and Claire McKay Bowen. 2024. "A Privacy-Preserving Validation Server Version 2.0." Washington, DC: Urban Institute.

About the Authors

Erika Tyagi is the associate director of data science products at the Urban Institute, where she develops scalable, impactful, and sustainable data science tools that advance innovative policy research. Her work includes building applications such as open data portals, interactive data tools, and privacy-enhancing technologies that safely expand access to confidential data.

Silke Taylor is the associate director of software engineering in the Technology and Data Science office at the Urban Institute. In her work, she focuses on providing advanced technological solutions in cloud computing, API development, and parallel computing. Her extensive experience also includes developing applications that meet federal security standards, such as FedRAMP.

Graham MacDonald is the chief information officer and vice president of technology and data science at the Urban Institute, where he leads the strategic planning and implementation of research, operations, and communications technology and works with staff across Urban to improve access to data, analytics tools, and innovative research methods.

Deena Tamaroff is a lead user experience designer at the Urban Institute. Her work involves the application of user research insights and user-centered design principles to the development of Urban Institute tools and products.

Claire McKay Bowen is a senior fellow in the Family and Financial Well-Being Division and leads the Data Governance and Privacy Practice Area at the Urban Institute. Her research focuses on developing technical and policy solutions to safely expand access to confidential data for advancing evidence-based policymaking. She is also interested in improving science communication and ensuring everyone is responsibly represented in data. In 2024, she became an American Statistical Association Fellow "for her significant contributions in the field of statistical data privacy, leadership activities in support of the profession, and commitment to mentoring the next generation of statisticians and data scientists." Further, she is also a council member of the Inter-university Consortium for Political and Social Research, an advisory board member of the Future of Privacy Forums, and she is an adjunct professor at Stonehill College.

Joshua Snoke is a statistician at RAND. He researches statistical data privacy, algorithmic-aided decision-making, and workforce development. He has expertise broadly in statistical data privacy and has published technical papers and policy reports on topics such as synthetic datasets, model estimation across multiple databases, and differentially private algorithms. He focuses mainly on evaluating and developing practical applications of privacy-preserving methodologies to identify technologies which increase access to administrative or survey data while maintaining privacy and confidentiality.

Acknowledgments

This tool was funded by the Statistics of Income Division of the IRS and the National Science Foundation National Center for Science and Engineering Statistics (NSF/NCSES) [49100422C0008]. Original funding and support for the initial research and version 1.0 prototype was provided by the Alfred P. Sloan Foundation [G-2022-17149], Arnold Ventures, the Microsoft SmartNoise Early Adopter Accelerator Program, and NSF/NCSES. We are grateful to all our funders, who made it possible for the Urban Institute to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We would like to thank our dedicated partners at the Statistics of Income Division of the IRS; especially Conrado Arroyo, Victoria Bryant, Kelly Dauberman, Amanda Eng, Chris Rexrode, and Mark Xu for their amazing support. We also thank our entire Safe Data Technology project team, consisting of Nikhita Airi, Leonard Burman, John Czajka, Jordan Hardenburgh, Lillian Hunter, Surachai Khitatrakun, Robert McClelland, Josh Miller, Gabe Morrison, Evy Park, Eden Phillips, Jeremy Seeman, Jean Clayton Seraphin, Joshua Snoke, Aaron R. Williams, and Doug Wissoker. Thank you to our group of testers who provided early feedback, including staff from the following agencies: Bureau of Justice Statistics, Bureau of Labor Statistics, Census Bureau, Labor and Workforce Development Agency at the State of California, National Center for Science and Engineering Statistics, and Nebraska Statewide Workforce & Educational Reporting System. Thank you also to Forum One and Graphicacy, for design and front-end development support in previous versions of the prototype. We finally thank our advisory board of key representatives and top experts across industry, academia, and government for their input in this work. The members are John Abowd, Jim Cilke, Connie Citro, Jason DeBacker, Rick Evans, Dan Feenberg, Max Ghenis, Nick Hart, Matt Jensen, Barry Johnson, Ithai Lurie, Ashwin Machanavajjhala, Shelly Martinez, Robert Moffitt, Amy O'Hara, Mauricio Ortiz, Nancy Potok, Jerry Reiter, Rolando Rodriguez, Emmanuel Saez, Wade Shen, Aleksandra Slavković, Salil Vadhan, and Lars Vilhuber.

The validation server version 3.0 prototype is a product of the Urban Institute's Safe Data Technologies Project. More information on the project is available at urban.org/projects/safe-data-technologies.

The authors welcome feedback on this paper. Please send all inquiries to validationserver@urban.org.

500 L'Enfant Plaza SW
Washington, DC 20024
www.urban.org

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization founded on one simple idea: To improve lives and strengthen communities, we need practices and policies that work. For more than 50 years, that has been our charge.
By equipping changemakers with evidence and solutions, together we can create a future where every person and community has the opportunity and power to thrive.