



CASE STUDY

How The Center for New Data Revolutionized Research on Sensitive Data with Immuta

2021

 IMMUTA™

The Center *for* New Data

The Customer

The Center for New Data is a nonprofit focused on accelerating research where new and novel data can provide valuable insights and open doors for ethical public problem solving.

The organization was initially established as the COVID Alliance in response to the pandemic and has evolved to tackle other critical policy issues through data. The Center relies on utilizing commercial, donated, and public-access data sets for analysis by their volunteer corps of data scientists and academic researchers. With access to this data, different classes of researchers are able to collaborate to produce replicable, modular research that can in turn be published in open-access journals and be used to drive social impact and non-partisan policy work. Its current initiatives include the COVID Alliance, focused on using data to inform a better pandemic response, and Observing Democracy, a non-partisan program aimed at using geolocation data to ensure equal access to voting.

The Challenge

The Center for New Data's mission of enabling ethical public problem solving is predicated upon its ability to provide researchers with new and novel data sets — most of which contain highly sensitive data.

For example, under its COVID Alliance program, the Center utilized a geolocation data set derived from 50 million Americans' mobile phones during the 2020 Thanksgiving holiday to understand the prevalence of social distancing during the holiday in an effort to control the spread of the virus. This and other sensitive data, however, can be

their collaboration with academic social science and economics researchers from universities and think tanks. Given this diverse and ever-changing group of contributors, the Center for New Data faces the complexity of needing to share sensitive data stored in a compartmentalized data lake (using Snowflake as the primary datastore) to

“ As we add more components in a cloud database environment, it's much cleaner than any sort of on-prem situation we've had before. The ability for us to manage access controls, deploy privacy enhancing technologies, and rapidly implement novel frameworks of governance for our research teams has been a breath of fresh air, with no management or overhead costs for adding additional cloud database solutions. ”

Ryan Naughton,
Co-Executive Director, Center for New Data

easily re-identified — an important concern for the Center due to legal, ethical, and contractual guidelines that protect against re-identification. It was critical for the organization to find a way to balance data privacy and utility through a mix of data controls, such as masking, and context controls, such as role- and attribute-based access controls.

A key component of the Center for New Data's approach to research and problem solving is bringing together diverse contributors from interdisciplinary backgrounds. Consequently, they recruit data science and data engineering volunteers from various industries and facilitate

many different researchers — all of whom need access to different data for varying purposes through the analytics environment. The process involved reviewing each research proposal and then creating new native views and roles in Snowflake to govern access to sensitive data for each research project. This approach quickly became too cumbersome and complicated, particularly in the context of a fast-spreading virus where real-time data was paramount. With limited resources, it was also critical for the process to be easily managed by the data engineers, who set up the analytics environments for the researchers.

The Solution

To deliver on its mission, the Center for New Data partnered with Immuta to build a data access governance layer that simplifies data sharing while preserving both data privacy and utility.

When researchers want to begin a new project, they can discover data sets through Immuta's active data catalog and workflows. Self-service workflows enable researchers to request access to data, acknowledge approved usage purposes, request access control changes, and propose new projects. This approach exponentially simplifies the data request process for the Center's data engineering team. New researchers now have instant access to data, but only the data required to work on their problem scope – reducing time to utilize data from months to hours.

In addition to removing barriers to data access, the Center for New Data leverages Immuta to automate the enforcement of fine-grained, attribute-based access controls (ABAC), and privacy enhancing technologies (PETs) on the data stored in Snowflake. Instead of relying on Snowflake controls, non-technical data governance experts can define robust access policies in Immuta. These policies are then natively enforced when users interact with data in Snowflake. This new approach greatly simplifies the process of preparing and sharing sensitive data for analytics, eliminating the complexity of creating roles for every researcher and building secure views and complex, role-based functions to implement advanced PETs like format-preserving masking or dynamic k-anonymization.

“ This purpose-driven access control system has been revolutionary to our efforts. Traditional role-based access controls often result in more and more grants over time, resulting in researchers having access to too much data beyond the scope of their current project. Instead, with Immuta, researchers’ access to data is governed by their project. When you’re in a project, you only have access to the specific data sources with the specific masking or anonymization policies in place necessary for you to conduct research for that project. This centralization of access controls and manageability through an intuitive web UI has empowered our data governance team to not rely upon database administrators to properly do their job. ”

Ryan Naughton,
Co-Executive Director,
Center for New Data

Results

By implementing Immuta, The Center for New Data has been able to:

- Save over \$1 million annually in data engineering costs
- Simplify the data request process for researchers, reducing the time to data by 30x, from 90 days to 3 days
- Eliminate the need to manage hundreds of individual data access policies by moving from RBAC to ABAC
- Reduce the number of access control policies to fewer than 10, which can be easily authored by non-technical data governance experts

Future Plans

While the Center for New Data primarily uses Snowflake today, they will soon add Databricks and Big Query to their research platform. With Immuta already in place, the Center will be able to govern data access across all their databases from Day One, enabling faster time to data access and more simplified security across multiple cloud data platforms.

