

## The NYU Administrative Data Research Facility (ADRF)

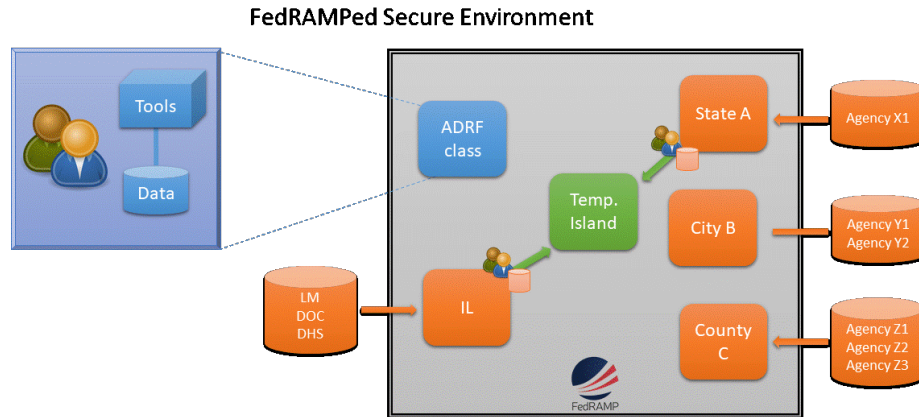
---

The data revolution is transforming how executives manage operations and businesses deliver goods and services. Yet when it comes to communities and the government agencies that serve them, the revolution has barely begun. The startup costs are high for creating an integrated data infrastructure about people served by different agencies and programs. Data are typically fragmented, and attempts at integration are often one-off and piecemeal. There are strong disincentives to share data across programs or with external parties. The NYU ADRF has developed technical and human approaches to enable community access to and use of data on human subjects. It built on many years of successful experience to design an infrastructure that was commissioned by the US Census Bureau to inform the decision making of the [Commission on Evidence Based Policy](#) It tackles three core issues.

- (i) Technical: it provides a secure environment – the Administrative Data Research Facility (ADRF) - within which data providers can place and share their data across agency and jurisdictional lines.
- (ii) Operational: it develops data facility administrative tools that permit disparate data to be found, documented and used – through discovery, stewardship and collaboration modules.
- (iii) Legal and practical: it engages the workforce – through Applied Data Analytics training programs - to ensure that there is a value associated with the data linkage that is both consistent with the agency mission and useful enough to engage decision-makers.

The approach has been extraordinarily successful. In the past year the initiative has provided services to over 200 government agency staff and researchers, and hosted almost 50 confidential government datasets from 12 different agencies at all levels of government. The approach was featured in the Commission report.

*Step 1: [Secure environment](#)* The ADRF institutionalizes secure access to and use of confidential data. It is a secure cloud-based environment that is [FedRAMP](#) certified and has received Authority to Operate from the US Census Bureau. The stamp of approval provides data owners with confidence that their data are secure. The cloud environment allows agencies within the same state or different states to agree to share their data in a common area in the cloud for specific approved projects.



*Step 2:* [Discovering and using data](#): Getting data from different agencies in the same place is necessary but not sufficient. The ADRF is developing tools that are inspired by Amazon and Trip Advisor. Our gamification approach enables users to contribute content about the data itself (variable definitions, coverage, limitations), preprocessing (code and lookup tables), analysis (code snippets and derived data), how the data have been used (research topics, methods and topics), who has used the data (names and locations) and code that has already been used to process and link data.

*Step 3:* [Creating value](#). The last step is building agency will to link data through training class. The pressures to meet existing program needs make it difficult for agencies to try something new and create pipelines of new products based on linked data. Government salary structures make it difficult to hire and retain enough in-house data analysts, so agencies don't have the capacity to work with new linked data. These combined challenges have led to the current situation - agencies cannot get the resources necessary to make use of new data, and because they don't use new data, they don't get new resources.

The ADRF approach has been critical to breaking this negative spiral. We have developed training classes that create a sandbox environment within which agency staff – not outside vendors – provide concrete evidence of the value of linking data. The approach has to be built around agency needs and use modular learning approaches. Our classes (i) create a pipeline of new product prototypes central to agency missions, (ii) develop teams of skilled practitioners who have the capacity to both link data and apply modern analytical approaches to cross agency problems, and (iii) make a growing set of linked data available as an ongoing asset for budget analysis and program management. The initial classes linked data from the Departments of Corrections (ex-offenders), Human Services (TANF and SNAP) and Housing (Section 8) with Department of Employment Security (wage records) to examine the effect

of neighborhood characteristics and access to jobs on the earnings and employment outcomes and subsequent recidivism and retention on welfare of ex-offenders and welfare recipients.

## Work with data across agencies

