Third Sector
capital partners

Secure Computing Needs in Pay for Success Projects:
Balancing Data Access and Privacy

30 August 2018

INTRODUCTION

At Third Sector, we partner with governments to think critically about how social services are contracted to enhance the achievement of positive life outcomes, shifting incentive structures from cost-reimbursement towards outcomes-oriented models, such as Pay for Success (PFS). Given this approach, many of our workstreams rely heavily on timely access to data in order to evaluate the unmet need across jurisdictions, identify the characteristics of the intended beneficiary population, and define outcomes that help us quantify impact for the community.

Over the past seven years, we have found that data availability is often the biggest limiting factor during project design and evaluation. Establishing data use agreements and ultimately transferring the data from the government sponsor to an independent evaluator can take up to a year, while the restrictions placed on the data use and scope as safeguards to privacy can undermine the usability of the dataset. As such, the efforts to secure data access creates three major impediments to projects:

I. Transactional costs for each one-off data use agreement
II. Need to forego valuable, yet restrictive data elements due to privacy constraints
III. Erosion of data value caused by significant lag times

Privacy. The legal protections applied to sharing someone's data. These privacy regimes are either established by the data owner or through guidance based on the type of data. For example, health data is subject to HIPAA requirements. Often the controls change based on whether the sharing include the full-field, cell-level info, or aggregated statistics. De-identified data is often easier to share than data that includes personally identifiable information, because it is considered non-private data.

Secure computing is a viable option to alleviate this tension between accessing information and protecting privacy on a secure system. Secure computing, in our usage, refers to any method of exposing data for analysis without providing direct access to the individual-level data. There are several approaches being developed or that have been deployed. Examples include remote analysis on secure data enclaves[1] where only analytical results

[1] Platt, Richard. "Data Enclaves for Clinical and Administrative Data Sharing." JAMA Internal Medicine, American Medical Association, 6 Aug. 2018, jamanetwork.com/journals/jama/fullarticle/2696616.

are passed back, homomorphic encryption[2] that allows for analysis without decrypting the data, or any other method where analysis can be done securely and reliably without sharing individual-level data that would necessitate informed consent and a data sharing agreement. Implementation of secure computing approaches would expand the use of government data without increasing the risk of privacy breaches, while also maintaining personal privacy. Secure computing holds the potential to address one of the biggest barriers to the implementation of outcomes-oriented contracting.

Security. The governance and technical protections to ensure that once data is shared it does not fall into the wrong hands, is re-identified, or gets used for unapproved purposes. Even when data meets privacy requirements, proper system security is needed to ensure that the data is not used for alternative purposes outside of the original scope.

OPPORTUNITIES

Third Sector's work often touches upon sensitive data, such as information related to incarceration, health, and child welfare. Based on our experience, there are practices that would facilitate effective and efficient access, use, and augmentation of federal data. Expanded secure data access coupled with privacy protection for those receiving services is crucial to increasing the impact of social services through real-time program adaptations and data-driven policy decisions to meet the dynamic needs of the individuals served.

There are two main challenges that we face. The first is to reduce the barrier to sharing data through approaches that structurally maintain privacy by providing access to only those fields approved to share. The second challenge is that the system also restricts the use of data is to approved purposes, there is a log of data access, and it is robust to unauthorized access. Often the use cases touch on both of these challenges. Methods that protect private information are likely to be more secure and more secure systems reduce the risk when sharing data.

The optimal secure computing solution would promote privacy by allowing the minimum set of data for a single purpose to be shared for a finite period of time, log of access and use, and be unusable for other purposes. There are many secure computing approaches that could provide these capabilities to government agencies. Three specific use cases for our work are aggregating and matching multi-agency data, population-level analysis, and determining outcomes for delivered services. Below we will elaborate on a few use cases, where the deployment of secure computing could accelerate the expansion of outcomes orientation at scale.

[2] Greenberg, Andy. "Hacker Lexicon: What Is Homomorphic Encryption?" Wired, Conde Nast, 3 June 2017, www.wired.com/2014/11/hacker-lexicon-homomorphic-encryption/.

Data Use Agreements

In one West Coast city, we are working with the local government to achieve better outcomes through its services designed to combat homelessness. In collaboration with the county and state data providers, we are using multi-level data to determine unmet needs, design impact goals, and monitor progress towards those goals. Given the sheer number of data sources and the intricacies in combining and matching data across datasets, we had to establish an elaborate and time-consuming framework to execute upon a multi-party data use agreement.

Given the underlying requirement to maximize the impact of the project, this complexity was a significant challenge to the project schedule and complicated coordination. If the individual-level data could have been accessed, matched, and analyzed without exposing the sensitive data then these challenges could have been avoided. However, without these capabilities, we had to triage the data access and narrow our focus on only the most critical data sources. We were not afforded the luxury of time, and as such had to act quickly on the available data rather than choosing to wait on datasets that could have added value but were not part of the minimum viable set. Unfortunately, this is not an isolated occurrence, we are constantly balancing the need to stay on schedule with the desire to maximize performance in our project designs.

Additionally, we have found that as the degree of data sensitivity increases, the longer it takes to get a data use agreement in place. Our projects are constructed to augment impact, which naturally leads us to gravitate towards areas of greatest need in a community – primarily underserved populations. Some examples of the data we are working to integrate are health care utilization, health care claims, reports of child abuse or neglect, homeless shelter stays, behavioral health utilization, wage and employment, and criminal justice involvement data.

These datasets individually have their own privacy protections, such as HIPAA, FERPA, or privacy act limits, which are collectively imposed on the combined dataset. Secure computing may allow the encrypted data to be analyzed, thereby removing the need to expose private information, reducing the risk of a data spill, maintaining privacy, and accelerating the project.

Beneficiary Population Analysis

In another project, our goal was to understand the needs and outcomes of veterans in San Diego. This project aimed to improve employment prospects and economic security for veterans with service-connected disabilities (SCDs) by providing intensive case management and wraparound supports to participating veterans. The critical questions in project design were what are the unmet needs within this population and what is the baseline set of outcomes for those who are eligible for Vocational Rehabilitation & Education (VR&E) program-funded training. However, due to the sensitivities attached to veterans' data, we were unsuccessful in gaining direct access to the Department of Veterans Affairs (VA)'s data and needed an alternative approach.

Our analysis partner, the American Institutes for Research (AIR), examined several data sources to determine whether the datasets are representative of the target population, including the American Community Survey, National Longitudinal Study of Youth 1997, and the Behavioral Risk Factor Surveillance System. Using these sources, we created a synthetic comparison pool of individuals by mimicking the VA's VR&E eligibility criteria across these databases. We looked to baseline potential outcomes within these data sources that covered employment and earning, poverty status, highest level of education, frequency of poor mental/physical health days, and frequency of alcohol/marijuana consumption. While this project did not move forward, the public data was able to fill the gap of not having access to the primary VA data. Despite the fact that the operational coordination issues ultimately stopped the project from moving beyond the feasibility assessment, the learnings from this project reinforce the idea that public, non-sensitive data can be just as valuable as private, individual-level data in informing better decision-making. The project also made us wish the we had such ease in accessing and analyzing private data securely.

Ideally, if it were possible to leverage secure computing rather than create a synthetic analogue of the authoritative data, the project may have moved more quickly and could have possibly avoided the operational issues that prevented it from moving forward. This example further highlights how the challenges associated with data access can directly impact a project's success. Exploration of secure computing approaches may provide insights on how to identify and better serve the needs of the beneficiary population more quickly.

Outcome Metrics

A typical social service program today follows a cost-reimbursement structure, where a government agency reimburses providers based on self-reported inputs, like staff hours dedicated, the number of client referrals, or a per-person fee for each participant. The effect is that much of the information governments collect is related to compliance rather than how effectively the population is being served. This creates incentives to increase staff hours, client referrals, and days of programming, while underutilizing opportunities to incentivize effective programs that create multi-year positive changes in the lives of those served. The most effective and least impactful programs are both reimbursed on their cost structure. Attention is paid to metrics that do not inform resource allocation decisions or hold the agencies accountable for creating meaningful life outcomes.

Many reform efforts have begun and have started to shift focus from inputs to outputs. Rather than reimbursing for costs or referrals, they provide a payment for certificate completion, skills attainment, or immediate outcomes like job placement. While these are good and needed changes, there is still room for improvement. For example, job training programs could also look at sustained employment beyond 12 months, wage increases year-over-year, or non-employment related outcomes. Ultimately, even outcomes indirectly connected to employment could factor into repayment as well. Examples include reduced recidivism rates for training programs reaching formerly incarcerated citizens, increased housing stability for housing insecure trainees, or increased connection to a primary care provider. We expect job placement to be an important pillar towards the path to self-sufficiency but current limitations related to agency data and funding structures prevent rewarding improvements centered around the whole person.

While compliance is a part of good governance, there is more to gain – across all stakeholders (funders and providers alike) – when incentive structures are designed to hold programs accountable to measurable impact on the lives of individuals served. These outcome measures would allow agencies to continuously evaluate providers on the progress made towards their shared goals, as it relates to achievement of life outcomes. To realize this vision, outcome data must be available, used to impact decisions, and be accessible across office boundaries.

The value of secure computing is maximized when multiple data sources are brought together to better understand those receiving benefits and the impact on their lives. Often time, eligibility for services is determined through factors that appear in one agency's systems while the impact often extends across different agencies purview. For example, child care services may be connected with enrollment in Temporary Assistance for Families with Dependent Children (Dept of Human Services), but child care may allow for a family to have a parent return to work (Dept of Revenue) at a job that provides health insurance (Medicaid utilization), and improve third-grade reading scores (Dept of Education). Only through multi-agency outcome analysis can the total social impact of a program be measured.

A concrete example of multi-agency data more completely capturing the value of services is in our Empowering Families projects. Third Sector is working with a national cohort of projects in five states and two counties that is focused on the implementation of two-generation[3] approaches to better coordinate services that touch different needs of the family and support improved outcomes for the child, parent, and family. These projects would also benefit from better access to wage data, as wage growth is a primary family outcome for measuring multi-generational economic mobility[4].

Moreover, since children and parental wellness is a critical precursor to economic stability and mobility, access to health outcome measures from Medicaid claim and payment data could support two-generation approaches. This data shows how connected families are to a primary care physician, demonstrate their cost of care, and allow for a relative focus on preventive versus acute care. Getting access to health data, especially children's health data, has been a very challenging prospect for projects not only due to the Family Educational Rights and Privacy Act (FERPA) but also the Health Insurance Portability and Accountability Act (HIPAA). Finding a way to make the data more available while protecting privacy rights is an important path to understanding which intervention models successfully break the intergenerational cycle of poverty we witness today.[5]

[3] Ascend at the Aspen Institute defines a two-generation approach as creating opportunities for and addressing needs of both children and the adults in their lives together. The approach recognizes that families come in all different shapes and sizes and that families define themselves. "Guiding Principles." Ascend at the Aspen Institute, http://ascend.aspeninstitute.org/two-generation/what-is-2gen/.

[4] Chetty et al., "The fading American dream: Trends in absolute income mobility since 1940," Science, 28 Apr 2017: Vol. 356, Issue 6336, pp. 398-406, DOI: 10.1126/science.aal4617

[5] ibid

POTENTIAL SOLUTIONS

Sensitive data can be used securely through new technological approaches. There are a growing number of industry tools that can balance these competing demands. For example, secure computing is improving quickly and can now perform basic analysis on encrypted data so that neither the algorithm nor the data needs to be exposed to get analytical results. Companies like Galois and Duality are extremely active in this space.

Other secure computing approaches commercially available today allow for field- or cell-level access restrictions, meaning that access can be given to part of the database without exposing the entire set or needing to create a trimmed database that is can be sent to partners. Through tools like Apache Accumulo, remote queries can be made without the need to share private data. It is critical to have user-based identity management as part of this system so that access is connected to that user's need. Identity management also supports logging of access and activity to verify if a privacy violation occurred. By using these tools, the sensitive data can be shared, a permanent log of data access created, and privacy maintained.

We must keep in mind that tools are only part of the solution. A cultural adaptation to use the full power of the tools rather than maintain the same practices needs to occur. Significant efforts are necessary to allay concerns about privacy risks. Along with the development and deployment of secure computing tools, there is a clear need for leadership on establishing regulatory guidance and norms around secure computing, so that its value can be made tangible to government agencies.

Multiple approaches to secure computing are advancing very rapidly. There are many secure computing approaches available and being developed today and even more new tools will be available in the near future. It is critical that governments modernize their data systems to improve government accountability and maximize the effectiveness of services. As they modernize, security and accessibility should be designed directly into the architecture and access controls. Only by incorporating the application of the data into the design can the system address these needs. Third Sector is ready to help our partners think through what is possible and evaluate options to realize that vision.

THIRD SECTOR OVERVIEW

Third Sector Capital Partners, Inc ("Third Sector"), a 501(c)3 nonprofit advisory firm, has a seven-year track record of collaborating deeply with communities to re-write how they contract for social services, re-aligning vast amounts of public resources to move the needle on social problems. We have engaged with more than 40 jurisdictions across the country in developing and deploying outcomes-oriented strategies to align resources with results for communities. We are strengthening the pipeline of state and local governments and service providers prepared to implement Pay for Success projects, as well as creating opportunities for communities to better measure and evaluate the success of their social service programs.