

Combined Data Analyses

Goal Leader(s): Mona Siddiqui, Office of the CTO, APG Goal Leader
and Rashida Dorsey, ASPE, Deputy Goal Leader

Overview

Goal Statement

- Increase combined analysis of disparate datasets in order to achieve better insights. By September 30, 2019 HHS will develop and implement an enterprise wide data governance model. This will enable more efficient and effective processes for sharing inter-agency data beyond a dataset's primary purpose. Two use cases of inter-agency data sharing through this process will be tested.

Challenges

- Siloes - HHS has rich data assets that are collected and maintained by separate agencies within the Department. In many instances, Operating Divisions may have specific authorities or mandates to collect data making joint collection or integration of data collection activities difficult.
- Statistical agencies and units must adhere to legal provisions (e.g. CIPSEA) that can inhibit data sharing. HHS programs and services can also be restricted from sharing data by legal provisions. For example, the National Database of New Hires (NDNH) is a valuable data asset, but it contains the highest degrees of protection.
- CMS, IHS and other Health Insurance Portability and Accountability Act (HIPAA) covered entities must adhere HIPAA regulations for the disclosure and use of covered health information
- Agencies within HHS each serve specific missions and developed systems overtime to address their unique needs. However, reexamining the underlying infrastructure can help identify gaps, areas for improvement, and efficiencies. Specifically, HHS needs to develop and implement an enterprise wide data governance strategy.

Overview

Opportunity

- Multi-sector issues—when an issue touches more than one agency or crosses into multi-disciplines than joint research and analyses occur and benefit everyone involved.
- Technology--advances in computing and IT infrastructure allow for secure sharing of information and enable collaborative work.
- Methodology—advances in statistics, machine learning, data mining, and geographic analysis enable disparate datasets to be combined to reveal new insights and find patterns.

Leadership

Core Team

- Mona Siddiqui, APG Goal Leader
- Rashida Dorsey, Deputy Goal Leader
- Reimagine HHS Leveraging the Power of HHS Data Core Team

Goal Structure & Strategies

- HHS agencies have experience conducting combined analyses and linking datasets together (both internal and external). However, these are often treated as “one-offs”, experimental, or exploratory rather than a tool to support the overall mission. Encouraging agency collaboration to conduct this work will require providing tools, devoting staff time, and establishing a framework that benefits all involved.
- An HHS wide data governance plan will establish ownership and rules for data assets and improve data quality and uniformity.
- A streamlined data use agreement infrastructure will enable more seamless interagency data sharing
- Developing an infrastructure that allows for secure and simple sharing of data assets will enable agencies and staff to collaborate on issues that affect multiple agencies such as the opioid epidemic.
- Comparing commonly used data architectures and data management tools to open source software and file formats could reveal more efficient ways to analyze and manipulate data assets. Adopting and using flexible open source solutions would also increase the Department’s ability to partner with the private sector.

Summary of Progress – FY18 Q1

The goal is new this year. However, HHS agencies have conducted combined analyses before with Departmental data (internal) and created linkages with other federal departments (external).

Data merging:

- SAMHSA created an interactive map showing MAT facilities and potentially underserved areas: https://www.samhsa.gov/data/mat_map
- ACF commissioned research on homelessness using data from disparate sources: https://www.acf.hhs.gov/sites/default/files/opre/opre_homefam_brief1_hhs_participation_benefits_programs_033016_b508.pdf
- ACF research on connecting TANF clients to stable employment: https://www.acf.hhs.gov/sites/default/files/opre/final_using_data_to_connect_tanf_clients_122_115_complete.pdf

Data asset linkage:

- NCHS linked several surveys to CMS Medicaid Data. <https://www.cdc.gov/nchs/data-linkage/medicaid.htm>
- NCHS partnered with HUD to link NHIS and NHANES to HUD administrative data to understand the connections between housing and health. <https://www.cdc.gov/nchs/data-linkage/hud.htm>

Key Milestones

- Development of Department-wide Data Use Agreements (DUAs)
- Develop framework for Department-wide Data Governance Plan
- Implement Enterprise wide data governance model
- Deploy two pilots to test the process and infrastructure

Key Indicators

- Note: The metric being used is the development and implementation of an enterprise wide data governance model. Progress will also be measured through the use of two pilots that will test the model.