Arlington County Homelessness Data Lake Proof of Concept (POC)
Scope Document

POC Title
Arlington County Homelessness Data Lake Proof of Concept

Challenges
Arlington County Virginia's Department of Human Services (DHS) oversees five operational divisions delivering services that promote a community of healthy, safe and economically secure children, adults and families. DHS also serves as the local homelessness continuum of care (CoC) developing strategies to assist people who are homeless or at risk of becoming homeless.

Per their annual Point-in-Time Survey, they have seen a 56 percent reduction in the number of homeless persons in Arlington, from 527 in 2008 to 232 in January 2017. They have 5 shelters – 2 for families, 2 for singles, and one for victims of domestic violence, and operate a hypothermia shelter during the winter months (November to March). Street homelessness has decreased due to a concerted effort to develop and assign permanent supportive housing units – apartments with rental subsidies and supportive services for individuals who are permanently disabled – either physically or mentally. They are committed to providing safe and affordable housing for vulnerable populations to live.

While DHS has a homelessness management information system (HMIS) provided through Social Solutions to track and maintain information about those receiving homelessness support, that system is not connected to other systems that contain information about other services delivered sometimes to the same individuals. Without the ability to see who they are serving through which programs, there is not enough reliable data to understand which programs or combinations of programs are most effective for which subpopulations of residents. Without that understanding, they have difficulty making data driven decisions about where to best invest limited resources to further eradicate homelessness from their community.

Summary of POC Proposal
We will demonstrate the capabilities and potential of using a data lake and entity resolution to enable DHS to derive an unduplicated list of individuals served by DHS across 3 data sources (HMIS, Cerner, and Visual Homes) to answer these questions:

1. Are homeless individuals transitioning to subsidized housing?
2. Are individuals in subsidized housing transitioning to homelessness?
3. Are individuals with behavioral health needs (Cerner data) transitioning to homelessness (HMIS data)?
4. Are homeless individuals served through the Treatment on Wheels (TOW) program remaining unconnected to outpatient services after they are housed?

POC Success Criteria

- The solution will provide a deduplicated count of individuals served per program included in the POC dataset.
- The AWS team will help DHS design reports to answer the questions above.

- DHS will be able to add other datasources after POC completion to scale across all DHS programs.
- AWS will provide a breakdown of the cost and options of scaling across all DHS programs.

POC Assumptions

- DHS will continue to own the data and the datalake
- AWS will guide DHS through POC setup and execution (virtual meetings primarily unless otherwise needed in person)
- The POC will utilize Novetta entity resolution capabilities to identify when an individual in the dataset appears to match another and to what degree of certainty.
- DHS will commit resources to be involved with the POC and work with Alistair Mclean on this engagement.
  DHS