# U.S. Department of Health and Human Services (HHS), Administration for Community Living (ACL): Data Restructuring Project

## Final Project Summary

## March 2019

# Purpose and Objectives

The purpose of the Administration for Community Living (ACL) Data Restructuring Project was to review existing ACL and Census data sets; identify practical, efficient, and statistically sound approaches for linking those data sets; and create an operational approach that will allow for linking additional data sets in the future. The project was conducted from September 29, 2017 to March 28, 2019 with support from aging and disability researchers, data scientists, and systems development staff from New Editions Consulting and IMPAQ International. The primary objectives of the project were to:

- Develop an approach and timeline for the project.
- Propose structure for curating data sets for cross-data set analysis.
- Test the proposed data structure and approach.
- Prepare a workplan and timeline for restructuring all data sets.

# Key Accomplishments

The project team produced a range of deliverables contributing to several key accomplishments of the project. The Project Accomplishments section provides a detailed description of project accomplishments for each objective.

- *Technical Advisory Group (TAG)*: Recruited a 10-member TAG consisting of aging and disability expert stakeholders from state and federal government, academia, and private companies. Presented five (5), 90-minute webinars to describe project activities, explain recommendations for data restructuring, and discuss/incorporate TAG member input.

- *Summary Data Structure Package*: Developed three Summary Data Structure Packages, one for each group of data sets included in the project (i.e., ACL aging data sets, Census data sets, ACL disability data sets). The documents describe each data set (i.e., purpose, population, data structure, feasibility of merging), appropriate linkages between datasets, and use cases illustrating cross-data set analyses. Each document is accompanied by a metadata table providing complete variable level details for each data set. Combined, the Summary Data Structure Packages total roughly 85-pages and cover 10,487 variables from 16 separate data sets.

- *Testing Phase*: Built a SQL testing database including a total of 57 data files from three-years (i.e., 2013-2015) of the ACL SPR and Title VI data sets. Demonstrated proof of concept for the new data structure through multiple levels of validation and association testing. Developed 13-page Testing Phase document, which provides detailed explanations of the testing approach and findings.

- *Data Restructuring Workplan and Timeline*: Developed 20-page Workplan and Timeline document describing key tasks, timeframes, personnel, and level of effort for a multi-year plan to expand ACL data linkages and build a next generation, web-based platform to provide ACL data users with new ways to navigate and analyze aging and disability data sets.

# Next Steps

ACL intends to build on the efforts of the Data Restructuring Project in several ways. These efforts may include continued improvement and preparation of the "back-end" data structure. For example, improving data documentation of ACL data sets, expanding the testing database to include new years and data sets,

aligning common measures across ACL data sets, and designing alternative ways for ACL data users to navigate aging and disability data sets other than the data set-specific search methods currently available on the AGing Integrated Database (AGID) platform.

Further development of the "back-end" will help advance ACL toward the vision for a new platform to replace AGID, including advanced and intuitive features that allow for meaningful cross-data set analyses. Future efforts to build this new "front-end" would include activities that span the full systems development life cycle. For example, defining all requirements, developing wireframes, building and testing the system, as well as delivering and maintaining/operating the system.

# Project Accomplishments

### Objective 1: Develop and Approach and Timeline for the Review of ACL Data Sets

| Project Deliverable | Description |
| --- | --- |
| Technical Advisory Group (TAG) | Recruited 10-member TAG for the purpose of obtaining stakeholder input throughout the course of the project. |
| First TAG Webinar | Held a 90-minute webinar on November 21, 2017. The project team provided TAG members descriptions of the data sets on AGID and a summary of interview data collected about AGID prior to the start of the contract. TAG members provided feedback on these materials and insight into the types of analyses they would like to be able to conduct with ACL's aging and Census data. |
| The Approach and Timeline for the Review of ACL Data Sets | 10-page document that describes the project team's plan and timeline for reviewing ACL aging and Census data sets hosted on AGID. |

### Objective 2: Propose Structure for Curating Data Sets for Cross-Data Set Analyses

Objective 2a. Review Aging Data Sets Hosted on AGID (i.e., SPR, Title VI, NSOAAP, and NORS)

| Project Deliverable | Description |
| --- | --- |
| Summary Data Structure Package Document – Aging Data Sets | 24-page document that describes the data set, appropriate linkages, use cases, and visual schema for each ACL aging data set hosted on AGID. |
| Metadata Table – Aging Data Sets | Excel file with 3,339 proposed variables and 14 variables to omit in new data structure. |
| Second TAG Webinar | Held a 90-minute webinar on June 5, 2018. The project team presented the new data structure for aging data sets with use cases to demonstrate examples of possible cross-data set analyses. TAG members asked questions and provided feedback. |

Objective 2b. Review Census Data Sets Hosted on AGID (i.e., ACS PUMS, ACS Special Tabulations 2011-2015, Population Estimates, Decennial Census 2010, and Decennial Census 2010 Special Tabulations)

| Project Deliverable | Description |
|---|---|
| Technical Advisory Group (TAG) | Recruited 10-member TAG for the purpose of obtaining stakeholder input throughout the course of the project. |
| Summary Data Structure Package Document – Census Data Sets | 22-page document that describes each data set and the recommendation to include ACS PUMS as the only Census data set for cross-data set analysis in the new data structure. Describes appropriate linkages, use cases, and visual schema ACS PUMS. |
| Metadata Table – Aging and Census Data Sets | Excel file with 4,996 proposed variables and 14 variables to omit in new data structure. |
| Third TAG webinar | Held a 90-minute webinar on August 23, 2018. The project team presented the new data structure for aging and Census data sets with use cases to demonstrate examples of possible cross-data set analyses, emphasizing Census data. TAG members asked questions and provided feedback. |

## Objective 2c. Review Selected ACL Disability Data Sets (i.e., AT, DD Councils, UCEDDs, PADD, PAAT, CIL, and ILS)

| Project Deliverable | Description |
|---|---|
| Summary Data Structure Package Document: Disability Data Sets | 39-page document that describes the data set, appropriate linkages, use cases, and visual schema for each ACL disability data set. |
| Metadata Table – Aging, Census, and Disability Data Sets | Excel file with 10,487 proposed variables and 14 variables to omit in new data structure. |
| Fourth TAG Webinar | Held a 90-minute webinar on November 16, 2018. The project team presented the new data structure for aging, Census, and disability data sets with use cases to demonstrate examples of possible cross-data set analyses, emphasizing disability data. TAG members asked questions and provided feedback. |

## Objective 3: Test the Proposed Structure and Approach

| Project Deliverable | Description |
|---|---|
| Technical Advisory Group (TAG) | Recruited 10-member TAG for the purpose of obtaining stakeholder input throughout the course of the project. |
| Testing Database | Built a Microsoft SQL Server database with SPR and Title VI data for 2013, 2014, and 2015. Performed validation testing to verify the integrity and accuracy of the data after importing it into Microsoft SQL Server. Performed association testing to verify that proposed technical and conceptual linkages between SPR and Title VI data are operational. |
| Testing Phase Document | 14-page document that provides detailed descriptions of the testing approach and findings. |

| Project Deliverable | Description |
| --- | --- |
| Fifth TAG Webinar | Held a 90-minute webinar on January 9, 2019 with ACL staff and one TAG member who served as a data science expert on the TAG. The project team demonstrated the testing database using a range of use cases. Participant asked questions and provided feedback. |

## Objectives 4 and 5: Prepare Workplan and Timeline for Restructuring all ACL Data Sets

| Project Deliverable | Description |
| --- | --- |
| Workplan and Timeline Document | 20-page document that describes the project team's recommended next steps to build on the efforts of the Data Restructuring Project. The document includes key tasks, timeframes, personnel, and level of effort for a multi-year plan to develop the "back-end" and "front-end" of a next generation, web-based platform to replace AGID. |