U.S. Department of Health and Human Services, Administration for Community Living: Data Restructuring II Project

Final Report
September 20, 2021

Table of Contents

Executive Summary

Introduction and Purpose

As the primary federal agency responsible for Older Americans Act (OAA) programs, the Administration for Community Living (ACL) is charged with the critical mission of supporting the Aging Services Network and the millions of older adults and individuals with disabilities who depend on it for their health, safety, well-being, and independence. In order to understand the reach, activities, and effectiveness of these programs, ACL manages and oversees annual program performance reporting by grantees and conducts additional data collection activities, such as large national surveys of aging and disability services programs. ACL also leverages external datasets, such as population datasets from the U.S. Census Bureau (Census), to capture more aging and disability data for its community of data users. This community of data users includes federal, state, and local program staff, policymakers, researchers, and other aging and disability stakeholders.

ACL makes these data available to its data users primarily through the AGing, Independence, and Disability (AGID) Program Data Portal. AGID provides a single, publicly-available, web-based platform for users to query a variety of ACL and external datasets, visualize and export results, and access additional resources. However, the current AGID system needs more modernized functionality to improve the user experience, more clear and comprehensive data documentation for some datasets, and a data structure that allows users to query across datasets.

The purpose of the ACL Data Restructuring II (DR II) project is to review existing ACL and Census datasets; identify practical, efficient, and statistically sound approaches for linking those datasets; create an operational approach that will allow for linking additional datasets in the future; and develop the foundational requirements for the next iteration of the AGID system (AGID 2.0). The project was conducted by ACL from September 26, 2019, to September 25, 2021, with support from aging and disability researchers, data scientists, and systems development staff from New Editions Consulting and IMPAQ International.

Navigating the Report

The ACL DR II project consists of four main tasks. Accordingly, this final report is organized into four sections, each corresponding to a project task. Each section describes the purpose of the task, approach, findings, and recommendations.

The report is modular: each of its sections can be read and understood separately from the other sections. Reading all sections of the report will provide the fullest picture of the ACL DR II project and accomplishments, but it is not necessary to read any one section of the report to understand another. For example, a reader who is most interested in learning about the Microsoft SQL database developed under the ACL DR II project may choose to only read Section 2: Data Repository for Aging and Disability Datasets. Occasionally, throughout the report, the reader is referred to another report section where a particular concept is defined and described in more detail. Additionally, each section includes references to appendices and supplemental materials. These materials provide further information for the reader that did not fit in the main body of the report or was otherwise not suitable for including in it.

A summary of each section of the report is provided below.

Section 1: Data Documentation

The purpose of the Data Documentation task was to develop detailed data documentation for selected aging and disability datasets. This section provides data documentation at the dataset level and variable level for seven datasets:

- four datasets for aging services grant programs authorized under the OAA: State Performance Report (SPR), Title VI, National Ombudsman Reporting System (NORS), National Survey of Older Americans Act Participants (NSOAAP);
- one dataset for population/demographic information from the U.S. Census Bureau: American Community Survey (ACS) Public Use Microdata Sample (PUMS); and
- two datasets for disability services grant programs authorized under the Rehabilitation Act, as amended by the Workforce Innovation and Opportunity Act of 2014: Centers for Independent Living (CIL) and Independent Living Services (ILS).

Dataset-level documentation consists of a narrative description of key characteristics of each dataset (e.g., data collection methodology, changes in the data collection over time, description of the underlying information management system and datafiles) and key considerations for linking datasets for valid and appropriate cross-dataset analysis. This section also provides a schema, or diagram, that illustrates a proposed data structure for relating all seven datasets included in the ACL DR II project. This schema serves as a blueprint for a new AGID database that enables cross-dataset analysis.

Variable-level documentation consists of a spreadsheet including all variables across all datasets included in the ACL DR II project and important information about each variable (e.g., variable description, variable type, data table and dataset where the variable is located, and applicable merge keys that can be used to relate the data for cross-dataset analysis).

Section 2: Data Repository for Aging and Disability Datasets

Under the Data Repository for Aging and Disability Datasets task, the ACL DR II project team built and tested a Microsoft SQL Server database following the data structure and schema developed through the Data Documentation task. The database integrates 30 datafiles from across all seven datasets, spanning a 9-year time frame (2010–2018). The key steps for developing the database included

- receiving the datafiles for all datasets and years,
- transforming the raw data into a standard file format,
- importing the data into the SQL Server Staging Database,
- processing and preparing the data for cross-dataset analysis and saving to the Enclave Database,
- creating a Common Database of data tables that list and define codes for key variables, and
- saving and transferring backups of SQL Server databases between ACL DR II project partner organizations.

The task also included validation and association testing of the database. The purpose of validation testing was to verify that the data maintained their integrity through the steps for importing the raw datafiles into SQL Server and preparing the data for analysis. The purpose of association testing was to verify that linkages for relating datasets were operational using real-world queries, or use cases, developed in collaboration with various ACL program leads and subject matter experts.

Results from the testing show that, overall, the data maintained their integrity through importing and processing in SQL Server. The results also show that linkages were operational. This section concludes with a summary of key lessons learned from database development and testing that will serve as important considerations during the development of AGID 2.0.

Section 3: Aligning Measures for Conceptual Linkages Across Datasets

The purpose of the Aligning Measures for Conceptual Linkages Across Datasets task was to identify linkages that exist between datasets included in the ACL DR II project. Linkages serve as the key for conducting valid cross-dataset analysis with ACL and U.S. Census Bureau (Census) datasets. There are two types of these linkages:

- Technical linkages exist between data that match across datasets on the basis of time and geography only.
- Conceptual linkages exist between data that match on the basis of time and geography, and also on the basis of subject area and measurement.

For example, an AGID user interested in OAA nutrition services might define and run their query to pull data on the number of home-delivered meals participants and the number of home-delivered meals served in their state over the past year. These data are technically linked between the SPR and Title VI datasets, since both are 1-year data collections available at the state level. These data are also conceptually linked because the measures are both about home-delivered meals and use exactly the same definitions in both datasets. Using this linkage, the AGID data user can examine the SPR and Title VI home-delivered meals data side by side as well as directly compare the data from the two datasets, for their state over the past year.

This section summarizes the availability of technical and conceptual linkages across datasets included in the ACL DR II project. It also describes strategies for increasing the number of conceptual linkages across ACL and external datasets over time, including

- leveraging existing approaches from related ACL projects to review and align measures across datasets (e.g., redesign of OAA grantee annual program performance report forms);
- considering federal standards and guidance to align measures on common topic areas (e.g., race, ethnicity, sex, primary language, and disability status);
- using periodic opportunities, such as U.S. Office of Management and Budget (OMB) approval renewals, to revise data collections for more closely aligned measures across datasets; and
- using ongoing mechanisms, such as the ACL Data Council, to review and align measures across datasets.

Section 4: Foundational Requirements for AGID 2.0

Under the Foundational Requirements for AGID 2.0 task, the ACL DR II project team performed business analysis, requirements gathering, and documentation activities to produce the foundational requirements document. These requirements are designed to guide ACL and their future development team through planning, analysis, building, testing, and launching AGID 2.0. This section summarizes key requirements materials developed by the ACL DR II project team, including

- business requirements which list and describe the functions for AGID 2.0 that ACL has identified to date, a priority level that indicates the importance of the business requirement, and any comments (e.g., notes, decision points, recommendations) for future consideration;
- user stories which translate the business requirements into a functional blueprint for AGID 2.0 through short, simple descriptions of how the user will interact with AGID features, told from the user perspective;
- process flows which visually depict the core functions and features of AGID 2.0 based on the business requirements and user stories; and
- proofs of concepts which illustrate the AGID 2.0 features that users will interact with and experience online, without representing the precise placement or style of the components of the feature (i.e., wireframes).

This section also describes next steps and recommendations for ACL and their future development team to further refine and prepare the foundational requirements, including

- addressing open items and tasks related to existing business requirements and identifying new business requirements;
- refining existing user stories and creating new user stories;
- testing and validating proofs of concepts; and
- reprioritizing business requirements to creating an AGID 2.0 Roadmap.

Conclusion

AGID is an important resource for publicly-available aging and disability data, but certain features and functions are lacking in the current system. The ACL DR II project team reviewed selected ACL and Census datasets, designed and validated a data structure for linking datasets and performing cross-dataset analysis, and developed the foundational requirements for AGID 2.0. This final report, in combination with the appendices and supplemental materials, summarizes this work and serves as a guide for ACL in planning, analysis, building, testing, and launching AGID 2.0. Key considerations and recommendations for ACL and their future development team are noted throughout the report and in the appendices and supplemental materials, where appropriate.

Section 1: Data Documentation

Introduction and Purpose

The purpose of the Data Documentation task for the ACL DR II project was to develop detailed data documentation for selected aging and disability datasets from ACL. Data documentation includes dataset-level and variable-level documentation. Dataset-level documentation consists of a narrative description of key characteristics of each dataset and considerations for linking each dataset for cross-dataset analysis. Variable-level documentation consists of a spreadsheet including all variables across the datasets in the ACL DR II project and important information about each variable. Dataset-level and variable-level documentation were designed to serve as resources for the development team who will build the data structure that connects ACL, Census, and other datasets in the future Aging, Independence, and Disability (AGID) 2.0 platform through which users will access these data.

Datasets and Documentation

The ACL DR II project includes seven datasets:

- four datasets for aging services grant programs authorized under the OAA: State Performance Report (SPR), Title VI, National Ombudsman Reporting System (NORS), National Survey of Older Americans Act Participants (NSOAAP);
- one dataset for population/demographic information from the U.S. Census Bureau: American Community Survey (ACS) Public Use Microdata Sample (PUMS); and
- two datasets for disability services grant programs authorized under the Rehabilitation Act, as amended by the Workforce Innovation and Opportunity Act of 2014: Centers for Independent Living (CIL) and Independent Living Services (ILS).

Each dataset is briefly described in the bullets below. More detailed information about each dataset is presented throughout this report and is available from the ACL website (https://acl.gov/) and the AGID Program Data Portal (https://agid.acl.gov/).

- **OAA Title III SPR:** The SPR is the primary source for U.S. states and territories to report annually on OAA Title III services that they provide (e.g., supportive services, nutrition services, caregiver services).
- **OAA Title VI Program Performance Report:** Title VI of the OAA authorizes grants to tribal and native organizations representing Older American Indians, Alaskan Natives, and Native Hawaiians. The Title VI dataset is the primary source for these grantees to report annually on OAA Title VI services that they provide (e.g., supportive services, nutrition services, caregiver services).
- **OAA Title VII NORS:** The NORS is the primary source for U.S. states and territories to report annually on services that they provide under Title VII (Chapter 2) of the OAA. The NORS dataset captures information on long-term care ombudsman efforts on behalf of residents in long-term care facilities.
- **NSOAAP:** The NSOAAP is a collection of six annual, national surveys of participants who received selected OAA Title III services (case management, transportation, congregate meals, home-

delivered meals, homemaker, and caregiver). The dataset includes information on perceptions of service quality, self-reported outcomes, demographics, physical and social functioning, and other service use.

- **ACS PUMS:** The ACS PUMS is a sample of the actual responses to the ACS, an ongoing Census survey that provides important information about social, housing, economic, and demographic characteristics of the U.S. population. The ACS PUMS data provide annual estimates for the U.S. population at the national and state level.
- **CIL:** The CIL dataset captures annual administrative reporting on independent living activities for 354 CIL grantees across the country, including information about their funding, staff and board members, individuals served, services and activities, and outcomes.
- **ILS:** The ILS dataset captures annual administrative reporting about independent living activities for U.S. states and territories, including information about funding, staff, individuals served, outcomes, community activities and collaborations, training and technical assistance needs, and progress independent living goals and objectives.

Dataset-Level Documentation

Dataset-level documentation is provided in this report for five datasets (SPR, Title VI, NORS, ILS, and CIL). Appendices A through E provide dataset-level documentation for each of these datasets. Each appendix is divided into sections that describe the underlying aging or disability program, data collection and methodology, changes in the data collection over time, data submission processes and information management systems, datafile structure, feasibility and validity of merging the data, and a schema for the dataset. Throughout each appendix, there are tables that provide lists and descriptions of additional supporting documentation and further resources for understanding each program and dataset, with links where applicable.

Although the ACL DR II project also includes the NSOAAP and ACS PUMS datasets, extensive dataset-level documentation already exists for these two datasets (e.g., survey methodology documents, survey reports, presentation slides and webinars, frequently asked questions [FAQs]). More information about this existing documentation for NSOAAP and ACS PUMS is discussed below.

Variable-Level Documentation

Variable-level documentation is provided for all seven datasets included in the ACL DR II project. This documentation can be found in the Excel workbook entitled ACL_DR_II_Variable_Level_Documentation.xlsx. The workbook includes seven worksheets, one for each dataset. Each worksheet includes a row for each variable in the dataset, along with a description of that variable, variable type (numeric, text, or currency), data table and dataset where the variable is located, and applicable merge keys (i.e., variables that can be used to relate the data for cross-dataset analysis).

Table 1 below provides summary information about each dataset and the data documentation produced under ACL DR II for each dataset.

Table 1: ACL DR II Datasets and Documentation

Dataset Name	Type of Data	Type of Dataset	Source	Dataset-Level Documentation	Variable-Level Documentation
SPR	Aging	Administrative	ACL	Yes	Yes

Dataset Name	Type of Data	Type of Dataset	Source	Dataset-Level Documentation	Variable-Level Documentation
Title VI	Aging	Administrative	ACL	Yes	Yes
NORS	Aging	Administrative	ACL	Yes	Yes
NSOAAP	Aging	Survey	ACL	No	Yes
ACS PUMS	Population	Survey	Census	No	Yes
ILS	Disability	Administrative	ACL	Yes	Yes
CIL	Disability	Administrative	ACL	Yes	Yes

Existing Dataset-Level Documentation: NSOAAP

The national, annual, cross-sectional survey of OAA Title III services recipients, NSOAAP, uses a complex, multistage sampling design. NSOAAP is structured as a collection of six surveys, each covering service recipients for a separate OAA Title III service type (case management, transportation, congregate meals, home-delivered meals, homemaker, or caregiver). NSOAAP is funded by ACL and administered by Westat. Each year, ACL and Westat produce a set of constructed datafiles for NSOAAP which are uploaded to the AGID Program Data Portal. AGID users are able to query the NSOAAP data by year, data element, stratifier, and geography using the Custom Tables feature in AGID. They are also able to download the raw NSOAAP datafiles for analysis outside the AGID platform.

Extensive data documentation to help understand and effectively navigate the NSOAAP already exists. Dataset-level documentation for NSOAAP is available from AGID and not provided in this report. Table 2 below describes existing data documentation that is available for NSOAAP and the location of the data documentation.

Table 2: Description and Location of Existing NSOAAP Data Documentation

Type of Data Documentation	Location of Data Documentation
Presentation slides for NSOAAP, covering the overall survey and each survey module, including an overview, data examples, and processes and procedures	AGID\Resources\Data Sources\ACL-Related Files\National Survey of Older Americans Act (OAA) Participants
Inventory of datafiles and data documentation for each year (2003–2005, 2008, 2009, 2011–2018) and survey module (case management, transportation, congregate meals, home-delivered meals, homemaker, caregiver) of NSOAAP. The following data documentation is provided for each year and survey module of NSOAAP: • codebook/frequencies • survey instrument • survey documentation • programming statements This data documentation is provided alongside downloadable datafiles for NSOAAP in multiple file formats.	AGID\Data Files\National Survey of OAA Participants

Existing Dataset-Level Documentation: ACS PUMS

The ACS is an ongoing Census survey that captures social, housing, economic, and demographic characteristics of the U.S. population. Over 3.5 million households across the U.S. are contacted each year to participate in the ACS. The ACS PUMS data contain an annual sample of the ACS records

representing about 1% of the population (e.g., of the U.S., of each state). Each year, ACL pays the U.S. Census Bureau to produce a custom datafile with annual estimates for the 60-and-older population, which are uploaded to the AGID Program Data Portal. These data include characteristics of older adults that are of specific interest to OAA programs (e.g., age, gender, race/ethnicity, education, marital status, disability, income, employment, household composition). AGID users can view the ACS PUMS data using the Data-at-a-Glance feature or query the ACS PUMS data by year, data element, stratifier, and geography using the Custom Tables feature in AGID.

Like NSOAAP, extensive data documentation already exists for ACS PUMS. Dataset-level documentation for ACS PUMS is available from the U.S. Census Bureau and not provided in this report. Table 3 below describes existing data documentation that is available for ACS PUMS and the location of the data documentation.

Table 3: Description and Location of Existing ACS PUMS Data Documentation

Type of Data Documentation	Location of Data Documentation
Main web page for the U.S. Census Bureau providing information about the ACS, from which, a visitor can navigate to information covering a range of topics for the ACS including: • general information and background on the ACS • survey design and methodology • data tables and tools • guidance for ACS data users (e.g., briefs, fact sheets) • technical documentation (e.g., variable lists and definitions)	U.S. Census Bureau, ACS web page: https://www.census.gov/programs-surveys/acs
Archive of American Community Survey questionnaires for each year of the survey (1996–2021)	U.S. Census Bureau, ACS Questionnaire Archive web page: https://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html
Main web page for the U.S. Census Bureau providing information about the PUMS files, a subsample of the ACS. From this main web page, a visitor can navigate to information covering a range of topics for the ACS PUMS: • general information and FAQs for the ACS PUMS • guidance on how to access ACS PUMS data • technical documentation (e.g., data dictionary, code lists)	U.S. Census Bureau, ACS PUMS web page: https://www.census.gov/programs-surveys/acs/microdata.html
Official, centralized, public platform for accessing data from the U.S. Census Bureau, including ACS PUMS data	U.S. Census Bureau, Data.census.gov website: https://data.census.gov

Data Structure and Relationships

AGID is an online query system that provides access to ACL-related aging and disability program performance data, survey data, and other data. The purpose of the system is to provide a single, user-friendly site for accessing aging and disability data from ACL and other sources (e.g., U.S. Census

Bureau). The system allows users to pull data through various pathways (data-at-a-glance, state profiles, custom tables, and datafiles) to produce and download data tables and visualizations.

In the current AGID system, users are able to define and run their queries on one dataset only. There is no data structure that specifies the relationships between datasets to permit queries of multiple datasets simultaneously. Defining this data structure and the relationships between datasets to support valid and appropriate cross-dataset analysis is a key objective of the ACL DR II project. At the core of this data structure for cross-dataset analysis are two types of data linkages: technical linkages and conceptual linkages.

Technical Linkages

Technical linkages exist between data that match across datasets on the basis of time and geography. For example, an AGID user might be interested in using technical linkages to produce a report with information about OAA nutrition programs in their state over the past year. In this scenario, the user might define and run their query to pull data on the number of home-delivered meals participants and the number of home-delivered meals served in their state over the past year. These data would come from the SPR and Title VI datasets. This query of multiple datasets is valid and appropriate because both SPR and Title VI provide data at the state level[1] for a 1-year period. The user can then examine these data side-by-side to understand the total number of home-delivered meals participants and meals served through both programs within the same geographic area and timeframe.

Exhibit 1 displays the technical linkages that exist among datasets reviewed as part of the ACL DR II project.

Exhibit 1: Technical Linkages Across Datasets

Dataset	Years[2]	Time Period of Datafiles	National	Census Region	ACL Region	Census Division	State	Title VI Grantee Level	CIL Grantee Level
SPR	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
Title VI	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-
NORS	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
NSOAAP	2011–2018	1 Year	**Yes**	**Yes**	-	-	-	-	-
ACS PUMS	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
CIL	2015–2017	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	**Yes**
ILS	2016–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-

[1] Title VI data are collected at the grantee (tribe or tribal consortium) level but can be aggregated to the state level.
[2] Does not include all years of available data for each dataset, only years of data included in the ACL DR II project.

Conceptual Linkages

Technical linkages are simpler and more common than conceptual linkages, since they only require that data match on the basis of time and geography. Conceptual linkages exist between data that match on the basis of time and geography but also on the basis of subject area and measurement.

Continuing the example above, home-delivered meals participants and meals served are both defined and measured in the same way in the SPR and Title VI datasets. Both datasets are administrative data collections that define home-delivered meals participants as the *unduplicated total number of eligible persons served one or more home-delivered meals*. Both datasets define home-delivered meals served (service units) as the *total number of meals served, where one meal is equivalent to a minimum of 33.3% of Dietary Reference Intakes (DRI) and complies with all other requirements of the OAA, state/local laws, and United States Department of Agriculture (USDA)/Department of Health and Human Services (HHS) Dietary Guidelines for Americans*.

Since these variables are conceptually linked across the datasets, the user can do more than examine these data side by side. For example, the user can create a ratio of the number of home-delivered meals served per participant to more directly compare how many meals the average participant received in the two programs. This comparison would not be valid and appropriate if the two variables were technically linked but defined and measured differently in the SPR and Title VI datasets.

In contrast, technical linkages allow the user to examine data side by side but not directly compare them. For example, a user may want to produce a report including a section with information on annual cases and complaints data from NORS for their state over the past 3 years. For context, the user may also want to include demographic trends among the older adult population in the state over the same 3-year period. Leveraging technical linkages, the user could define and run their query to return cases and complaints information from NORS and demographic information from ACS PUMS for their state over the past 3 years. This technical linkage is valid and appropriate because NORS and ACS PUMS provide 1-year datafiles with estimates at the state level. However, since NORS does not measure demographic information and ACS PUMS does not measure cases and complaints information, there is no conceptual linkage between these two datasets on these measures.

Since conceptual linkages have greater requirements than technical linkages, they are less common and do not exist between all datasets. Exhibit 2 shows which pairs of datasets, among all the datasets included in the ACL DR II project, have any variables that are conceptually linked. The name of each dataset is listed twice in the exhibit, once in the top row and once in the first column. The diagonal gray line fill pattern represents the intersection of a dataset with itself or a duplicate intersection of two different datasets. These table cells should be ignored. To determine the availability of conceptual linkages for one dataset with all others included in the ACL DR II project, be sure to review the row and column information for the dataset of interest. For example, the fourth row indicates that NORS has conceptual linkages with SPR and Title VI. The fourth column indicates that NORS does not have conceptual linkages with NSOAAP, ACS PUMS, CIL, and ILS. The availability of conceptual linkages for NORS with all other datasets can only be understood by reviewing the fourth row and fourth column.

Exhibit 2: Conceptual Linkages Across Datasets

	SPR	Title VI	NORS	NSOAAP	ACS PUMS	CIL	ILS
SPR							
Title VI	**Yes**						
NORS	**Yes**	**Yes**					
NSOAAP	**Yes**	**Yes**	**–**				
ACS PUMS	**Yes**	**–**	**–**	**Yes**			
CIL	**Yes**	**Yes**	**–**	**–**	**Yes**		
ILS	**Yes**	**Yes**	**–**	**–**	**Yes**	**Yes**	

More information about conceptual linkages is provided in Section 3: Align Measures for Conceptual Linkages Across ACL Datasets, including the specific variables that are conceptually linked across pairs of datasets.

Overall Schema

All seven datasets included in the ACL DR II project can be related for cross-dataset analysis following the data structure shown in Exhibit 3. This exhibit provides a schema, or diagram, that illustrates the contents of all seven datasets and the relationships between the datasets. It serves as a blueprint for a database that follows the rules defined by technical and conceptual linkages to relate the datasets for cross-dataset analysis. This database was developed under the ACL DR II project and is discussed in Section 2: Data Repository for Aging and Disability Datasets.

The overall schema is divided into two major columns. The left column illustrates how the seven datasets can be related using technical linkages. All seven datasets are stacked in this column and connected by through a vertical bar with the variable names GeoID (geographic identifier) and Year. This structure illustrates that all seven datasets can be technically linked on the basis of having data that share a common geography and time. Using the example described above, the nutrition services query would select the data user's state using GeoID, the past year using Year, and the total numbers of home-delivered meals participants and home-delivered meals using NumericValue for those two variables in both the SPR and Title VI datasets.

For easy reference, each dataset and variable is numbered. The number and order assigned to the datasets and variables have no specific meaning. The numbers are added only for ease of communication during the database development process. They provide a standard and simple way to account for and communicate about all the datasets and variables included in the database. Also, the numbering can easily scale to include additional datasets and variables that may be added in the future. Like the numbering used in the overall schema, colors have no specific meaning.

The right column provides a summary of the conceptual linkages that exist between datasets. For each dataset it shows the topic areas with variable(s) that can be conceptually linked with at least one other

dataset. For example, under 6.0 CIL, the fifth topic listed is Improved Access. The schema shows that CIL can be conceptually linked with ILS on one or more variables related to improved access to certain types of services for consumers (e.g., transportation, health care services, assistive technology). The specific variables that can be conceptually linked between datasets are not shown in the overall schema due to space limitations. The next section of this report provides a schema for each individual dataset, including more information on conceptual linkages. Section 3: Align Measures for Conceptual Linkages Across ACL Datasets provides detailed explanation and discussion of conceptual linkages across datasets.

Exhibit 3: Overall Schema

Technical Linkages

Linkages between data that match across datasets on the basis of time (Year) and geography (GeoID).

GeoID, Year

1.0 SPR

1.1 GeoID, 1.2 Year

1.3 Org	1.8 FieldLabel
1.4 Component	1.9 Type
1.5 ACLRefNumber	1.10 NumericValue
1.6 SectionName	1.11 CurrencyValue
1.7 Field	1.12 NonNumericValue

2.0 Title VI

2.1 GeoID, 2.2 Year

2.3 Org	2.8 FieldLabel
2.4 Component	2.9 Type
2.5 ACLRefNumber	2.10 NumericValue
2.6 SectionName	2.11 CurrencyValue
2.7 Field	2.12 NonNumericValue

3.0 NORS

3.1 GeoID, 3.2 Year

3.3 Org	3.8 FieldLabel
3.4 Component	3.9 Type
3.5 ACLRefNumber	3.10 NumericValue
3.6 SectionName	3.11 CurrencyValue
3.7 Field	3.12 NonNumericValue

4.0 NSOAAP

4.1 GeoID, 4.2 Year

4.3 J_YRSRVC_STRAT_VAR_ID
4.4 J_YRSRVCGRPVAR_VAL_SortOrder
4.5 J_YRSRVCGRP_VAR_ID
4.6 J_YRSRVCGRP_VAR_SortOrder
4.7 VARLABEL_ID
4.8 VARLABEL_Label
4.9 VARVAL_ID
4.10 VARVAL_Formatted
4.11 STRAT_ID
4.12 STRAT_Description
4.13 J_YRSRVCSTRATCATSTRAT_VAL_ID
4.14 J_YRSRVCSTRATCATSTRAT_VAL_SrtOrd
4.15 STRATVAL_ID
4.16 STRATVAL_Formatted
4.17 J_YRSRVCSTRATCAT_STRAT_ID
4.18 J_YRSRVCSTRAT_SortOrder
4.19 J_YRSRVCSTRATCAT_STRAT_SortOrder
4.20 SERVICE_Description
4.21 Variable
4.22 Value

5.0 ACS PUMS

5.1 GeoID, 5.2 Year

5.3 AGEGRP_ID	5.6 Field
5.4 POP_ID	5.7 Value
5.5 SEX_ID	

6.0 CIL

6.1 GeoID, 6.2 Year

6.3 GrantNumber	6.6 Datatype
6.4 NameofCenter	6.7 Numeric_Value
6.5 Field	6.8 Non_Numeric_Value

7.0 ILS

7.1 GeoID, 7.2 Year

7.3 GrantNumber	7.6 Datatype
7.4 Field	7.7 Numeric_Value
7.5 FieldLabel	7.8 Non_Numeric_Value

Conceptual Linkages

Linkages between data that match on the basis of time and geography, but also on the basis of subject area and measurement.

1.0 SPR

Topic	Dataset(s)
Staffing	Title VI
Total Expenditures	NORS
Nutrition Services Persons Served	Title VI, NSOAAP
Supportive Services Persons Served	Title VI, NSOAAP
Caregiver Services Persons Served	Title VI, NSOAAP
Nutrition Services Units	Title VI
Supportive Services Units	Title VI
Caregiver Services Units	Title VI
Demographics	ACS PUMS, CIL, ILS, NSOAAP
Living Alone	ACS PUMS, NSOAAP
Total Cases/Clients Served	CIL, ILS
Functional Limitations	NSOAAP

2.0 Title VI

Topic	Dataset(s)
Staffing	SPR
Total Expenditures	NORS
Nutrition Services Persons Served	SPR, NSOAAP
Supportive Services Persons Served	SPR, NSOAAP
Caregiver Services Persons Served	SPR
Nutrition Services Units	SPR
Supportive Services Units	SPR, NSOAAP
Caregiver Services Units	SPR

3.0 NORS

Topic	Dataset(s)
Total Expenditures	SPR, Title VI

4.0 NSOAAP

Topic	Dataset(s)
Demographics	SPR, ACS PUMS
Functional Limitations	SPR, ACS PUMS
Nutrition Services Persons Served	SPR, Title VI
Supportive Services Persons Served	SPR, Title VI
Caregiver Services Persons Served	SPR
Living Alone	SPR, ACS PUMS
Supportive Services Units	Title VI

5.0 ACS PUMS

Topic	Dataset(s)
Demographics	SPR, NSOAAP, CIL, ILS
Living Alone	SPR, NSOAAP
Functional Limitations	NSOAAP

6.0 CIL

Topic	Dataset(s)
Demographics	SPR, ACS PUMS, ILS
Total Cases/Clients Served	SPR, ILS
Service Type	ILS
Increased Independence	ILS
Improved Access	ILS
Cases Closed	ILS
Funding	ILS
Persons With Disabilities Served	ILS

7.0 ILS

Topic	Dataset(s)
Demographics	SPR, ACS PUMS, CIL
Total Cases/Clients Served	SPR, CIL
Service Type	CIL
Increased Independence	CIL
Improved Access	CIL
Cases Closed	CIL
Funding	CIL
Persons With Disabilities Served	CIL

Section 2: Data Repository for Aging and Disability Datasets

Introduction and Purpose

The purpose of the Data Repository for Aging and Disability Datasets task was to expand the Microsoft SQL Server database developed under the ACL Data Restructuring I (DR I) project to include additional years of data and additional datasets. This expanded SQL Server database follows the data structure and relationships described in Section 1: Data Documentation. Table 4 below shows the datasets and years of data included in the SQL Server database for the ACL DR I project and the expanded SQL Server database for the ACL DR II project. Data are included from the following datasets: SPR, Title VI, NORS, NSOAAP, ACS PUMS, CIL, and ILS.

Table 4: ACL DR I and II SQL Server Databases

Dataset Name	Included in ACL DR I SQL Server Database	Years of Data Included for ACL DR I	Included in ACL DR II Expanded SQL Server Database	Years of Data Included for ACL DR II
SPR	Yes	2013–2015	Yes	2010–2018
Title VI	Yes	2013–2015	Yes	2010–2018
NORS	-	-	Yes	2009–2018
NSOAAP	-	-	Yes	2011–2018
ACS PUMS	-	-	Yes	2010–2018
CIL	-	-	Yes	2015–2017
ILS	-	-	Yes	2016–2018

This task also included testing the expanded SQL Server database for both validation and data association. The purpose of validation testing was to verify that the data imported into the expanded SQL Server database matches the raw data files. The purpose of association testing was to demonstrate that technical and conceptual linkages between datasets are operational.

This section describes the technical requirements and steps used to build the expanded SQL Server database and perform validity and association testing. It also describes the testing results and modifications made to the expanded SQL Server database to resolve the issues that emerged during testing. Finally, this section provides an overall discussion of the expanded SQL Server database development and testing phases and key considerations for the future development of AGID 2.0.

The expanded SQL Server database for the ACL DR II project is provided in the SQL file entitled ACL_DR_II_SQL_Database.sql.

Database Development

This task was performed in two major phases, the database development phase and the testing phase. The database development phase included setting up systems and infrastructure requirements to build the expanded SQL Server database, receiving datafiles from ACL for the selected datasets and years of data included in the ACL DR II project, reviewing and processing the datafiles, transforming the data (where appropriate), and importing the data into SQL Server for testing. Each of these steps is described in detail below.

Systems and Infrastructure Requirements

This task was performed using Microsoft SQL Server, a relational database management system. Additional tools were used to handle certain steps in the database development and testing phases. Table 5 below lists each of these tools, describes its purpose, and specifies whether or not the tool was used in the database development phase and testing phase.

Table 5: ACL DR I and II SQL Server Databases

System and Infrastructure Requirement	Purpose	Used in Database Development Phase	Used in Testing Phase
SQL Server 2008 R2 database server	Datafile storage in database	Yes	Yes
SQL Server Management Studio	Data validation and testing	Yes	Yes
R Version 4.0.4; RStudio Version 1.4.1106; R libraries	Data processing and importing into SQL Server	Yes	-
Secure file transfer protocol	Sharing large datafiles between ACL, New Editions, and IMPAQ	Yes	-
Microsoft Excel	Output file format for reviewing data validation results	Yes	-

The SQL and R programming files containing the code for data processing and database development and testing procedures are provided as supplemental materials to this report. The R programming file is entitled ACL_DR_II_R_Code_for_Data_Transformation,_Importing,_and_Validation_Testing.7z and the SQL programming file is entitled ACL_DR_II_SQL_Code_for_Association_Testing.sql.

Process for Receiving Datafiles from ACL

ACL provided a total of 30 datafiles to New Editions/IMPAQ for processing and importing into the expanded SQL Server database. The datafiles were provided in either .csv or .xlsx format. The combined datafile sizes ranged from 2.0 megabyte (MB) to 312.0 MB for each dataset. The source of the datafiles also varied by dataset. Datafiles for SPR, Title VI, and NORS were downloaded directly from the Older Americans Act Performance System (OAAPS). Datafiles from NSOAAP and ACS PUMS were provided by ACL contractors who constructed them from the underlying NSOAAP and ACS PUMS datasets. The CIL and ILS datafiles were downloaded directly from the AGID Program Data Portal. Table 6 below provides further details about the datafiles, by dataset.

Table 6: Datafiles from ACL

Dataset	Number of Datafiles	Format of Datafiles	Combined Size of Datafiles	Source of Datafiles
SPR	2	.csv	70.4 MB	ACL-OAAPS
Title VI	2	.csv	12.2 MB	ACL-OAAPS
NORS	2	.csv	2.30 MB	ACL-OAAPS
NSOAAP	6	.csv	312.0 MB	ACL-Contractor
ACS PUMS	12	.csv	88.9 MB	ACL-Contractor
CIL	3	.xslx	2.0 MB	AGID
ILS	3	.xslx	2.0 MB	AGID

When New Editions/IMPAQ received each datafile, it was downloaded and the datafile size was verified to ensure that the download process did not truncate it. Each datafile was reviewed to assess it for general completeness and appropriateness of the data. If we identified any issues with a datafile, we communicated the issues to ACL, and ACL provided the correct datafile.

Data Transformation

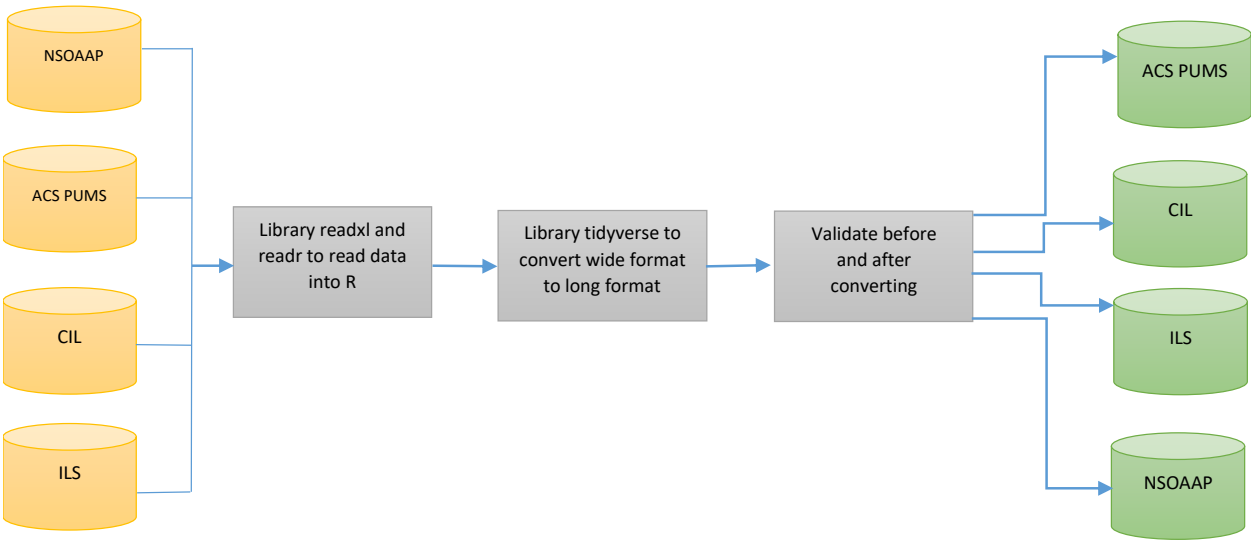
The datafiles we received for processing and importing into the expanded SQL Server database were in different datafile formats. The SPR, Title VI, and NORS datafiles were long formatted, meaning that each row in the datafile represented one data point for one variable, for a specific reporting year and state. Long formatting results in fewer variables (or columns) in the datafile and more rows. NSOAAP, ACS PUMS, CIL, and ILS datafiles were wide formatted, meaning that each row in the datafile represented all data points for all variables, for a specific reporting year and geographic area (such as state). Wide formatting results in more variables (or columns) in the datafile and fewer rows.

Consistent formatting across all datasets in the expanded SQL database is technically necessary to permit cross-dataset analysis. To be consistent with the OAAPS formatting style, we followed the steps below to transform wide data format into long data format:

1. Using R library readr, we read the four datasets into R environment.
2. We used R library tidyverse to remove leading or trailing blanks.
3. We used R library tidyverse/eshape2 to transform the data into a long format.

We conducted validity testing, described later in the report, before and after the transformation to ensure that data maintained integrity throughout the transformation process. Exhibit 4 illustrates the flow of the process for transforming the data from wide to long format.

Exhibit 4: Data Transformation Process



Color Code:
Yellow = Not ready for import to SQL server
Green = Ready for import into SQL server

Data Importing

Once the data transformations were completed and all datafiles followed a consistent format, the datafiles were imported into Microsoft SQL Server. Data importing occurred in two phases. First, the datafiles were imported into a SQL Server Staging Database. Then, the data were processed to prepare for cross-dataset analysis and saved to an Enclave Database.

The New Editions/IMPAQ team also created a Common Database containing data tables, such as the Geography table, used to elaborate on the datasets. For example, each source data table contains only the GeoID column. The Common Database provides more information about GeoID, such as the state name and state abbreviation for state-level GeoIDs. The Common Database serves as a data dictionary that can be used to explain coded variables.

A secure file transfer protocol was used to transfer backups of SQL Server databases between IMPAQ and New Editions. This way, if the database files were corrupted, or if there were other issues with the database, the backup files could be used to re-create the databases.

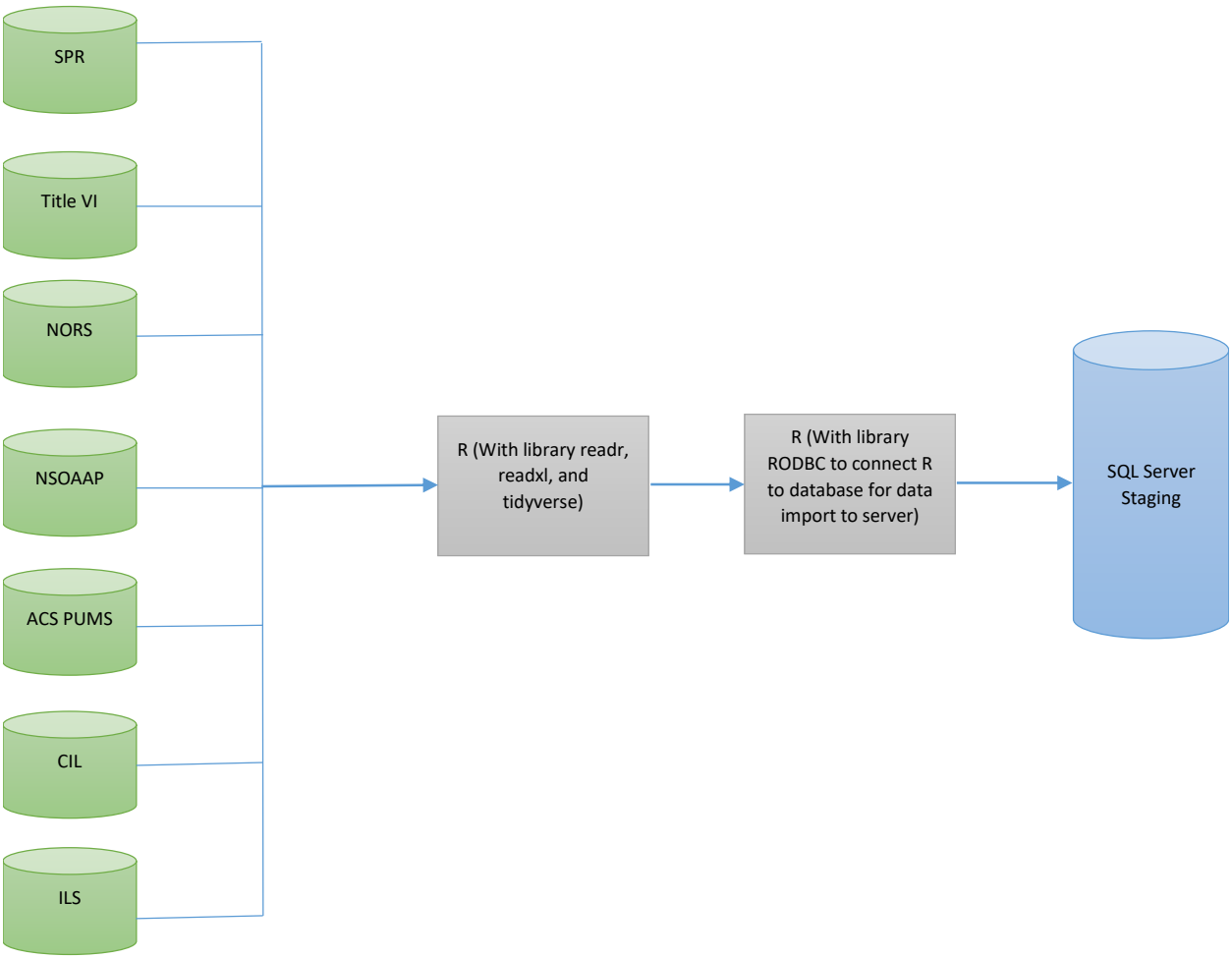
A technical explanation and illustration of the flow of data through the Staging Database and Enclave Database is provided below.

Staging Database

The following bullets explain the steps used to populate the Staging Database. Exhibit 5 provides an illustration of the steps. The staging database serves as storage for the raw data files before processing for final import to the Enclave Database.

- All datafiles received were formatted as .csv/xlsx datafiles. The R statistical package including the libraries readr, readxl, and tidyverse was used to read and process the data for importing/loading into the database.
- After the datafiles were read in the R statistical package and processed (that is, cleaned, formatted, and transformed for ACS PUMS, NSOAAP, ILS, and CIL), the R package RODBC was used to load the datafiles from the R environment into SQL Server database.

Exhibit 5: Populating the Staging Database



Enclave Database

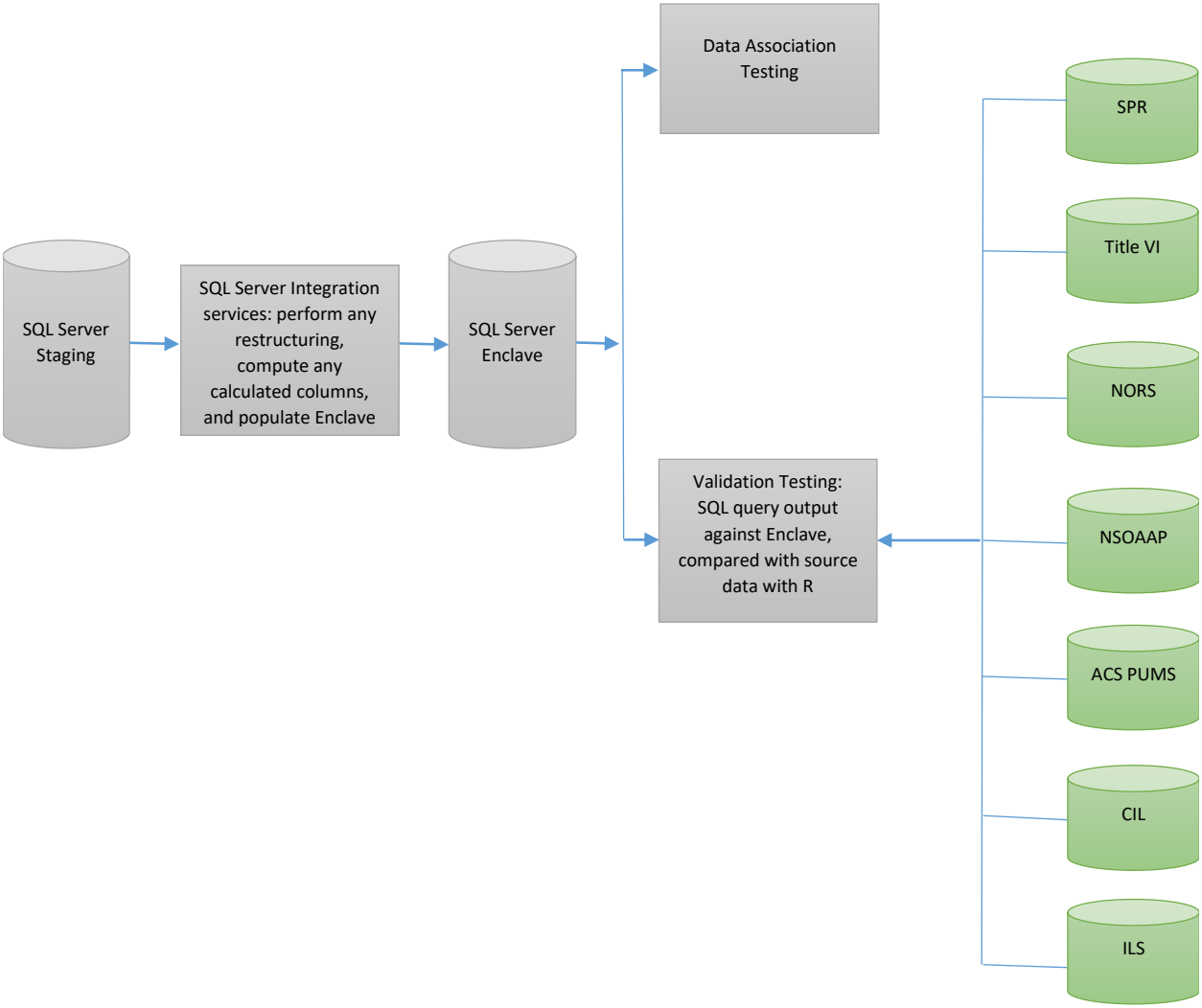
The following bullets describe the steps used for populating the Enclave Database. The Enclave Database is where the final data tables for all datasets—that is, the final files to be delivered to ACL and to be used for cross-dataset analysis—are stored. Exhibit 6 provides an illustration of the steps.

- Drop destination tables.
    - Destination tables are the target tables into which the data are loaded. If the target table name already exists, then it is deleted; otherwise, SQL Server will not accept the new table name.
- Create destination tables.
    - The primary key of each destination table was determined through data review by ACL and New Editions/IMPAQ. The primary key indicates which columns/variables make each row in the data unique.
- Populate destination tables, while

    - combining (stacking or appending vertically) data tables for the same dataset across multiple years;
    - combining (merging or appending horizontally) data tables to standardize the geographic identifier across datasets—a process of adding the GeoID and its description to each table to allow for cross-dataset analysis; and
    - computing calculated columns/variables (e.g., for newly constructed variables).
- Create association tables and populate them, as appropriate.
    - Since technical linkages are very common across datasets by state (and geographic levels that can be aggregated up from state, such as national) and by year, this association is not populated across datasets. Rather, an association table was created that documents which data tables can be technically linked by state and year.
    - Conceptual linkages are populated at the data table and column/variable levels for each column/variable in one table that is conceptually linked with a column/variable in another table.

A business key is a column used to link one or more data tables. The basic business keys in the Enclave Database are GeoID and Year. However, when the business key for two data tables in the Enclave Database differ, a new data table is created. For example, there may be a desire to compare demographic characteristics in CIL and ACS PUMS. In this case, the team would create a new data table keying on state, year, and demographic characteristics.

Exhibit 6: Populating the Enclave Database



Testing Phase

The purpose of the testing phase was twofold. First, the contents of the Enclave Database were compared to the source datafiles to ensure the data maintained integrity through data importing and processing. Then a series of use cases was developed for cross-dataset analysis, with input from ACL, and the team developed and ran programming code to demonstrate that the technical and conceptual linkages required for those use cases were operational. The use cases were designed to represent real-world queries of multiple ACL and Census datasets that ACL staff members and other stakeholders would run for their analysis and reporting purposes. Whenever test results indicated an error or other data issue, the issue was resolved, processes were modified to incorporate the solution, the test was rerun to verify the issue was resolved, and the issue and resolution were logged in a tracking table.

Approach and Timeline

During the ACL DR I project, the New Editions/IMPAQ team selected a relational database approach for completing the testing phase. A relational database is a type of database that stores and provides access

to data points that are related to one another. The advantage of relational databases is that they enable users to easily categorize and store data that can later be queried and filtered to extract specific information for reports. For consistency, the team continued to use this approach to test linkages across the seven datasets in the expanded SQL Server database. The approach is designed to be flexible, and it can easily accommodate additional datasets that may be added in the future.

Validation Testing

The New Editions/IMPAQ team conducted validation testing with all datasets by comparing numeric and currency variable-specific aggregates in the Enclave Database and source datafiles. Eight aggregates were used for validation testing. Table 7 lists each aggregate, provides a brief description, and gives the R code used to calculate the aggregate.

Table 7: Aggregates for Validation Testing

Aggregate	Description	R code
Sum	Returns the total sum of all numeric values for each variable	sum(col,na.rm=TRUE)
Min	Returns the minimum value for each numeric variable	min(col, na.rm=TRUE)
Max	Returns the maximum value for each numeric variable	max(col, na.rm=TRUE)
Avg	Returns the average of all numeric values for each variable	mean(col, na.rm =TRUE)
Count	Returns the total number of observations (rows) for each variable	length(col)
Count distinct	Returns the total number of distinct or unique values for each variable	length(unique(col[!is.na(col)]))
Count null	Returns the total number of null or missing values for each variable	length(col) - length(col[!is.na(col)])
Count non-null	Returns the total number of non-null or other-than-missing values for each variable	length(col[!is.na(col)])

The team also conducted a second stage of validation testing with the NSOAAP and ACS PUMS data because after the wide-formatted data had initially been processed and imported the into the Enclave Database, it was later decided to adopt consistent formatting across datasets by transforming NSOAAP and ACS PUMS data from wide to long. This change necessitated a second stage of validation testing after importing the long-formatted NSOAAP and ACS PUMS data into the database. The team conducted testing on the long-formatted data to ensure that the expected column and row counts were the same before and after the transformation. The same aggregates listed in Table 7 were applied for this second stage of validation testing. Table 8 provides a brief description of the transformation steps for these datasets and lists the number of records before and after the transformations.

Table 8: Data Transformations

Dataset	Description of Transformation	Total Number of Records *Before* Transformation	Total Number of Records *After* Transformation	Expected Number of Rows After Transformation
NSOAAP	Maintained all non-numeric fields and transposed all numeric fields from columns to rows. There was a total of 5 numeric fields transposed for each of the 6 data files.	1,595,902	7,979,510	7,979,510
ACS PUMS	Maintained all non-numeric fields and transposed all numeric fields from columns to rows. There was a total of 272 numeric fields transposed across the 12 data files.	287,271	6,759,320	6,759,320
CIL	Maintained the Grant Number, State, and Year fields and transposed all other columns to rows.	Example Table 1,347 rows and 21 columns	8,321 rows and 6 columns	8,321 rows and 6 columns
ILS	Maintained the Grant Number, State, and Year fields and transposed all other columns to rows.	Example Table 1,347 rows and 21 columns	18,882 rows and 6 columns	18,882 rows and 6 columns

The New Editions/IMPAQ performed validation testing for all datasets included in the ACL DR II project from February 3, 2020, to July 13, 2020. Table 9 lists each dataset, information about its source file(s), and the timeframe of its validation testing.

Table 9: Datasets for Validation Testing

Dataset	Number of Source Datafiles	Format of Source Datafiles	Combined Size of Source Datafiles	Total Number of Variables/ Columns for Validation Testing	Timeframe for First Stage of Validation Testing	Timeframe for Second Stage of Validation Testing
SPR	2	.csv	70.4 MB	1,644	2/17/20–2/21/20	N/A
Title VI	2	.csv	12.2 MB	33	3/16/20–3/20/20	N/A
NORS	2	.csv	2.30 MB	36	6/11/20–6/12/20	N/A
NSOAAP	6	.csv	312.0 MB	96	2/16/20–2/21/20	3/6/20–3/11/20
ACS PUMS	12	.csv	88.9 MB	331	2/3/20–2/10/20	3/4/20–3/6/20
CIL	3	.xslx	2MB	355	7/1/20–7/5/20	N/A
ILS	3	.xslx	2MB	808	7/10/20–7/13/20	N/A

Summary of Results

The New Editions/IMPAQ team did not identify any major differences or issues in comparing the aggregates between the Enclave Database and source datafiles for all datasets. The results indicate the data maintained integrity through importing and processing of source datafiles into the Enclave Database.

During validation testing, seven minor issues were identified and resolved across datasets. Table 10 lists and provides information on each issue, including when it was identified and when resolved, what the issue was and how it was resolved, and whether it was a processing or data issue. Processing issues include issues with data procedures used for validation testing, such as errors or omissions in programming code. Data issues include issues with the underlying dataset used in the validation testing.

Table 10: Validation Testing Results – Issues and Resolutions

Issue	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
1	SPR	2/18/20	2/21/20	Validation tests completed on "FieldLabel," which can be duplicative across sections of the SPR form.	Reran validation tests on unique ID "ACLRefNumber." Provided concatenated "ACLRefNumber," "FieldLabel," and value type (i.e., "NUMERIC," "CURRENCY," "TEXT") for variable/column in detailed validation report.	Processing
2	SPR	2/20/20	2/21/20	Differences in formatting decimal places in source datafiles and SQL produce very small/negligible differences in summation and average measure for 796 rows. Difference ranges from 0.0000000117 to 0.0000000003725.	None, negligible differences shown in detailed validation report Excel workbook.	Processing

Issue	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
3	SPR	2/20/20	2/21/20	Missing aggregate results for variables formatted as "CurrencyValue."	Determined "CurrencyValue" to be type of numeric data. Applied validation tests to these variables and updated report.	Processing
4	All	2/20/20	2/21/20	Initial R coding does not compute aggregates on variables with null or missing values.	Revised and reran validation tests using R coding that treats null or missing values as zeros.	Processing
5	ACS PUMS; NSOAAP	3/4/20	3/6/20	SQL Server is not efficient for loading large datasets. Transforming the data to match OAAPS datafile formatting requires extensive processing time.	Applied recommended approach to load the data in smaller chunks (e.g., partitioned NSOAAP homemaker data table into 30 chunks, each partition having less than 40,000 rows). In the future, data transformation and loading with large datasets in SQL Server should use the data partitioning approach and allow adequate processing time. This should be less of a concern when adding only single years of data to the expanded SQL database.	Processing

Issue	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
6	ACS PUMS	3/6/20	3/10/20	Differences in formatting decimal places in source datafiles and SQL Server produce very small/ negligible differences in summation and average aggregates for 48 rows. The range of these differences is from 0.00000000001.8 to 0.000000000009.9.	None, negligible differences shown in detailed validation report Excel workbook.	Processing
8	Title VI	3/16/20	3/18/20	Four tribe names in the Title VI Grantee Master List did not appear in the OAAPS Title VI datafiles (i.e., San Juan Pueblo, Stillaguamish Tribe of Indians, Yakutat Tlingit Tribe & Craig Community Association, Kuskokwim Native Association).	ACL advised on 3/18/20 that these four tribe names should not appear in the Title VI data due to a name change (San Juan Pueblo) or due to no longer being a grantee (Stillaguamish Tribe of Indians, Yakutat Tlingit Tribe & Craig Community Association, Kuskokwim Native Association) because former grantees are not included in OAAPS Title VI data.	Data

Issue	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
9	CIL	4/16/20	4/16/20	Differences in formatting decimal places in source datafiles and SQL Server produce very small/ negligible differences in summation and average aggregates for 13 rows. The range of these differences is from --0.00000000002.3 to 0.0000000002.3.	None, negligible differences shown in detailed validation report Excel workbook.	Processing
10	ILS	4/20/20	4/20/20	Differences in formatting decimal places in source datafiles and SQL Server produce very small/ negligible differences in summation and average aggregates for 14 rows. The range of these differences is from --0.0000005.9 to 0.0000001.1.	None, negligible differences shown in detailed validation report Excel workbook.	Processing

Detailed Validation Report

For detailed results of the validation testing for all datasets, see the Excel workbook entitled ACL_DR_II_Validation_Testing_Report.xlsx. This workbook includes worksheets that provide variable-level results of the validation testing for each dataset. There are two worksheets each for NSOAAP, ACS PUMS, CIL, and ILS, one for each stage of validation testing.

These detailed results include side-by-side, variable-specific aggregates for the source datafiles and Enclave Database data. For each aggregate, an additional column is provided with a formula calculating the difference between the results of the aggregate in the source datafile and Enclave Database data. All aggregates, for all variables, are listed as having a difference of zero or very small percentage (due to rounding or formatting differences across the two sets of data) to show that the data match. Green color highlighting is applied to provide a quick visual representation that the data match.

Association Testing

The purpose of association testing is to demonstrate that technical and conceptual linkages between datasets are operational. Association testing involved querying the Enclave Database employing use cases developed by ACL. Table 11 lists the 23 use cases that were employed to test technical and conceptual linkages across the seven datasets.

Approach and Timeline

Through discussions with ACL staff and subject matter experts in the ACL program areas, 23 use cases were identified for association testing. We identified the appropriate tables in the expanded SQL database for these use cases, then developed SQL code to query the Enclave Database for each use case. Some of the use cases were based on existing reports. In such cases, results were validated by comparing them to the reports.

SQL Code

The SQL code for association testing was written in Microsoft SQL standard, meaning the SQL Server engine supports all functions used in the code. This also means that the queries may not work if the SQL code is used in another SQL platform, such as PostgreSQL or MYSQL, that does not support all SQL Server functions. The SQL code developed for association testing does not require other software programs to run. Each use case is designed to run independently, so there are no dependencies.

Additionally, the SQL code for each use case defines the data table(s) used for the query, includes information on recoding of columns/variables, and includes annotations describing the purpose and details of the use case. The results of each query can be exported into Excel for ease of viewing. The SQL code is provided in the SQL file entitled ACL_DR_II_SQL_Code_for_Association_Testing.sql.

Table 11: Use Cases for Association Testing

Number	Description	Dataset(s)
1	Examine total number of congregate meals and home-delivered meals served between 2010 and 2018 in both the OAA Title III and Title VI programs in Oklahoma.	SPR and Title VI
2	Examine the number of congregate meals and home-delivered meals per client between 2010 and 2018 in both the OAA Title III and Title VI programs in Oklahoma.	SPR and Title VI
3	Examine the expenditures per OAA Title III congregate meal and per home-delivered meal between 2010 to 2018 in states in ACL Region III.	SPR
4	Determine if OAA Title III recipients were older than the general U.S. population of adults over age 60 in 2015.	SPR and ACS PUMS
5	Determine if participants who received OAA Title III home-delivered meals services in ACL Region 10 were older than the general population of adults ages 60 and over in ACL Region 10 in 2015.	SPR and ACS PUMS
6	Compare the annual percentage of OAA Title III case management participants living in poverty with the annual percentage of the general population of adults age 60 and over in Florida living in poverty from 2010 to 2018.	SPR and ACS PUMS

Number	Description	Dataset(s)
7	Compare the age distribution (percentage) of individuals served by the CIL program and ILS program across the U.S. in 2016 and 2017.	CIL and ILS
8	Compare the gender distribution (percentage) of individuals served by the ILS program in ACL Region 10 with the general population in ACL Region 10 in 2017.	ILS and ACS PUMS
9	Examine the number of OAA Title III caregiver services participants and the percentages of quality ratings (i.e., excellent, very good, good, fair, poor).	SPR and NSOAAP
10	Examine the number of OAA Title III case management service units provided per client and the length of time (i.e., 6 months to 1 year, 1–2 years, 2–5 years, 5 or more years) distribution (percentage) that case management services participants have received services.	SPR and NSOAAP
11	Compare the percentage of OAA Title III clients served who live in rural areas to the percentage of all adults ages 60 and over in the U.S. who live in rural areas in 2015.	SPR and ACS PUMS
12	Compare the percentage of OAA Title VI clients who had incomes below the poverty level to the percentage of all Native American adults ages 60 and over in the U.S. in 2017.	SPR and ACS PUMS
13	Examine differences in the percentage of CIL and ILS consumers who achieved their goals related to increased independence in the significant life area of self-advocacy/self-empowerment.	CIL and ILS
14	Compare the race/ethnicity distribution (percentage) of CIL participants with the general U.S. population of adults.	CIL and ACS PUMS
15	Examine the change in the number of OAA Title III home-delivered meals served from 2011 to 2018 and participant satisfaction.	SPR and NSOAAP
16	Compare annual changes in the number of full-time and part-time nutrition and supportive services staff members for all Title VI grantees in Wisconsin from 2012 to 2016.	Title VI
17	Determine the percentage of adults ages 60 and over in California from 2010 to 2018 who received OAA Title III home-delivered meals.	SPR and ACS PUMS
18	Compare the percentage of OAA Title III home-delivered meals clients in Illinois who identify as Black/African American with the percentage of adults ages 60 and over in Illinois who identify as Black/African American from 2010 to 2018.	SPR and ACS PUMS
19	Compare the percentage of OAA Title III funds spent on transportation in Illinois and Indiana in 2018.	SPR
20	Compare annual percentage of congregate meals participants who live alone with percentage that report their social opportunities have increased since becoming involved with the service in the U.S. from 2010 to 2018.	SPR and NSOAAP

Number	Description	Dataset(s)
21	Compare annual percentage of home-delivered meals participants who report difficulties with 3 or more activities of daily living (ADLs) or instrumental activities of daily living (IADLs) in the U.S. from 2010 to 2018.	SPR
22	List chronic conditions reported by home-delivered meals participants in the U.S. in 2018 in order from most to least commonly reported.	NSOAAP
23	Provide the percentage of home-delivered meals participants reporting each count of total number of chronic conditions (i.e., 1–16) in the U.S. in 2018.	NSOAAP

Summary of Results

The New Editions/IMPAQ team conducted association testing for small groups of two to four use cases at a time, then presented the SQL code and results to ACL during biweekly project meetings. The association testing occurred from August 5, 2020, to November 20, 2020. During this time, we verified that the technical and conceptual linkages were operational among the datasets and found minor issues with the data.

Database Updates from Association Testing

During association testing, the team identified and resolved minor issues with the datasets and data tables. Each issue is listed and described in Table 12. Since the testing focused on specific data tables that mapped to the use cases, all issues for these data tables were resolved, and the team performed a full scan of the Enclave Database to identify related data tables and apply the fix. After the scan, it was determined that there was a need to update the geographic identifier field for all data tables and add their descriptors. Geographic identifiers were renamed to be consistent across the database. For example, the geographic identifier field for SPR was Org, not GeoID. All geographic identifier fields were updated in the Enclave Database to be GeoID.

Table 12: Association Testing Results – Issues and Resolutions

Issue No	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
1	ACS PUMS	10/5/2020	10/6/2020	Linking variable in different format from other datasets; that is, GeoID formatted as numbers in Enclave Database but formatted as state abbreviation in other datasets.	Recoded GeoID to match with other datasets.	Data

Issue No	Dataset	Date Identified	Date Resolved	Issue	Resolution	Processing or Data
2	ILS	10/20/2020	10/20/2020	Age grouping labels differ from CIL dataset.	Recoded ILS age grouping labels to match CIL dataset.	Data
3	CIL	10/21/2020	10/21/2020	CIL years differ from ILS years.	Reformatted CIL years to be consistent with ILS years.	Data
4	ACS PUMS	10/22/2020	10/22/2020	Gender format labels differ from ILS.	Recoded gender labels in ACS PUMS.	Data
5	ACS PUMS	10/30/2020	10/30/2020	Race format labels differ from CIL.	Recoded race labels in ACS PUMS.	Data

Conclusions

Overall, the processes for developing and testing the database were implemented effectively and successfully, with minor issues. The team observed that loading large files into the database was time consuming; instead, partitioning extensive data into smaller parts to import was the best practice, as it was less time consuming. Also, transforming data for NSOAAP, ACS PUMS, CIL, and ILS datafiles from wide to long format and standardizing geographic names and other labels across the seven datasets were key factors to successfully building and demonstrating the database.

For the development of AGID 2.0, the following additional key points should be noted to the development team:

- There are no constraints on updating the data tables; that is, no rules were specified for the data in a table. Constraints are used to limit the type of data that can go into a data table.
- The data tables are long formatted. In the long format, each row is a one-time point per subject. So, each subject will have data in multiple rows. Any variables that do not change across time will have the same value in all the rows.
- The common variables/columns to link tables are GeoID and Year, which are available for each data table.
- The descriptions/labels for the columns/variables GeoID, ADL_ID, AGE_GRP_ID, IADL_ID, POVERTY_ID, and SERVICE_ID are saved in the Common Database.

Section 3: Aligning Measures for Conceptual Linkages Across Datasets

Introduction and Purpose

The purpose of the Aligning Measures for Conceptual Linkages Across Datasets task was to identify linkages that exist between selected aging and disability datasets from ACL. The linkages serve as the key for conducting valid cross-dataset analysis. Additionally, the task identified strategies to further align measures across datasets to increase the number of linkages for valid cross-dataset analysis in the future.

There are two types of linkages for cross-dataset analysis: technical linkages and conceptual linkages. These linkages are defined and described in Section 1: Data Documentation. They are revisited below with some additional details.

Technical Linkages

Technical linkages exist between data that match across datasets on the basis of time and geography. For example, an AGID user might be interested in using technical linkages to produce a report with information about OAA nutrition programs in their state over the past year. In this scenario, the user might define and run their query to pull data on the number of home-delivered meals participants and the number of home-delivered meals served in their state over the past year. These data would come from the SPR and Title VI datasets. This query of multiple datasets is valid and appropriate because both SPR and Title VI provide data at the state level[3] for a 1-year period. The user can then examine these data side by side to understand the total numbers of home-delivered meals participants and meals served through both programs within the same geographic area and timeframe.

Exhibit 7 displays the technical linkages that exist among datasets reviewed as part of the ACL DR II project.

Exhibit 7: Technical Linkages Across Datasets

Dataset	Years[4]	Time Period of Datafiles	National	Census Region	ACL Region	Census Division	State	Title VI Grantee Level	CIL Grantee Level
SPR	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
Title VI	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-
NORS	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
NSOAAP	2011–2018	1 Year	**Yes**	**Yes**	-	-	-	-	-
ACS PUMS	2010–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-
CIL	2015–2017	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	**Yes**
ILS	2016–2018	1 Year	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	-	-

All seven datasets collect 1-year data and can be technically linked with each other on the basis of 1 or more common years of data. Additionally, all seven datasets collect data that are representative at the national and Census Region geographic levels. The NSOAAP is the most limited of the seven datasets in

[3] Title VI data are collected at the grantee (tribe or tribal consortium) level but can be aggregated to the state level.
[4] Does not include all years of available data for each dataset, only years of data included in the ACL DR II project.

terms of technical linkages on the basis of geography. These data cannot be linked with data from other datasets at the ACL Region, Census Division, state, or substate geographic levels. The Title VI and CIL datasets are the only ones with data that are representative at a substate geographic level. These data can be aggregated to the state level, or higher geographic levels, and then technically linked with other datasets. In the example described above, the AGID user's query would sum the data on the number of home-delivered meals participants and the number of home-delivered meals served by each Title VI grantee within their state over the past year. These aggregated Title VI data could then be technically linked to the SPR data for these measures in their state over the past year.

Conceptual Linkages

Technical linkages are simpler and more common than conceptual linkages because they only require that data match on the basis of time and geography. Conceptual linkages exist between data that match on the basis of time and geography, but also on the basis of subject area and measurement.

To be conceptually linked, data from the datasets must be

- collected within comparable time frames (e.g., years),
- representative of the same geographic area (e.g., national level, state level),
- collected using related measures (e.g., matching questions, definitions of terms, allowable values or response options), and
- comparable with regard to other important parameters (e.g., method of data collection, such as administrative reporting or client self-report).

Continuing the example from above, data on both home-delivered meals participants and meals served can also be conceptually linked between the SPR and Title VI datasets.

SPR data cover an annual reporting period of October 1 to September 30, and Title VI data cover an annual reporting period of April 1 to March 31. Although these timeframes are not an exact match, they are comparable 12-month data collection periods that allow for valid conceptual linkage. Still, any differences in data collection timeframes should be noted to AGID data users so they may determine if the timeframes are comparable enough for the purposes of their specific analyses. In this example, if a data user requires more precisely matching data collection timeframes, they may choose not to conceptually link data from SPR and Title VI for their purposes.

SPR data are provided at the state level by OAA Title III state grantees. Title VI data are provided at the tribe or tribal consortium level by OAA Title VI grantees. These data do not represent the same geographic area; however, it is valid to aggregate all Title VI data for grantees within a state. Once the Title VI data have been aggregated to the state level, then the SPR and Title VI data represent the same geographic area and can be validly conceptually linked.

Both the SPR and Title VI datasets are administrative data collections. Grantees for both programs annually gather these administrative data using a standardized, federally approved program performance reporting data collection tool for their grant program. Additionally, both data collection tools define home-delivered meals participants as the unduplicated total number of eligible persons served one or more home-delivered meals. Both datasets also define home-delivered meals served (service units) as the total number of meals served, where one meal is equivalent to a minimum of

33.3% of DRI and complies with all other requirements of the OAA, state/local laws, and USDA/HHS Dietary Guidelines for Americans.

Since these data fulfill the criteria for conceptual linkages listed in the bullets above, they can be validly conceptually linked across the datasets. With technically linked data, the AGID data user can examine the data side by side. Conceptual linkages allow the AGID data user to go a step further and directly compare data across datasets. For example, the AGID data user can create a ratio of the number of home-delivered meals served per participant in the state to compare how many meals the average participant received in the two programs. This comparison would not be valid and appropriate if the two variables were technically linked but defined and measured differently in the SPR and Title VI datasets.

Since conceptual linkages have greater requirements than technical linkages, they are less common and do not exist between all datasets. Exhibit 8 shows which pairs of datasets, among all the datasets included in the ACL DR II project, have any variables that are conceptually linked.

Exhibit 8: Conceptual Linkages Across Datasets

	SPR	Title VI	NORS	NSOAAP	ACS PUMS	CIL	ILS
SPR							
Title VI	**Yes**						
NORS	**Yes**	**Yes**					
NSOAAP	**Yes**	**Yes**	**–**				
ACS PUMS	**Yes**	**–**	**–**	**Yes***			
CIL	**Yes**	**Yes**	**–**	**–**	**Yes***		
ILS	**Yes**	**Yes**	**–**	**–**	**Yes***	**Yes***	

*Indicates that the data collection tool for both datasets in the pair did not change because the conceptual linkages review was completed during the ACL DR I project. Conceptual linkages for these pairs of datasets were carried over from ACL DR I project final deliverables because the data collection tools have not changed and the conceptual linkages determinations from the ACL DR I project continue to apply.

Approach

During the ACL DR I and II projects, the New Editions/IMPAQ team conducted detailed reviews of pairs of datasets in collaboration with ACL to determine how to validly and appropriately link these datasets for cross-dataset analysis. The team identified key variables for technically and conceptually linking data across datasets and documented instances of potential linkages.

Potential linkages occur when two datasets include a similar variable but that variable isn't defined and measured in exactly the same way for both datasets. In some cases, it was determined with ACL that these potential linkages can be used to link data across datasets, but that requires a notice to the AGID data user explaining the differences in how the two variables are defined and measured in the two datasets. The AGID data user could then determine if the two variables are comparable enough to link

the data for the purposes of their specific analyses. If an AGID data user required more precise matching, they could choose not to link the data using these variables.

The team documented all linkages information in separate Excel spreadsheets, one spreadsheet for each pair of datasets. These spreadsheets are located in the Excel workbook entitled ACL_DR_II_Aligning_Measures_Spreadsheets.xlsx. Each spreadsheet includes detailed information for each linking variable from each dataset. Strategies for improving linkages across datasets are also proposed toward the end of the report.

The work with ACL to review datasets for technical and conceptual linkages has been a multistage process. During the ACL DR I project, the team worked with ACL to determine the technical linkages that serve as the fundamental keys to linking the datasets for cross-dataset analysis. The team also completed a first stage of dataset reviews to identify actual and potential conceptual linkages. Since completing the ACL DR I project, ACL revised the annual program performance report form for SPR, Title VI, and NORS. ACL documented detailed information describing changes to variables from the legacy to current versions of all three data collections.

During the ACL DR II project, the team reviewed these changes for all affected pairs of datasets and updated the conceptual linkages information. Conceptual linkages information was carried over from ACL DR I project final deliverables for pairs of datasets that were not affected as the data collection tools for these datasets did not change and the conceptual linkages determinations from the ACL DR I project continue to apply.

Our processes for determining technical and conceptual linkages on the ACL DR I and II projects are described below.

Aligning Measures Reviews - ACL DR I

Under the ACL DR I project, the team convened a 10-member Technical Advisory Group (TAG) consisting of aging and disability expert stakeholders from state and federal government, academia, and private companies. In collaboration with the TAG and a core team of ACL staff members, selected datasets were reviewed to determine linkages that could be used to validly and appropriately relate these data for cross-dataset analysis. Reviews of the data and documentation were performed to determine valid and appropriate merge rules following three steps:

- Step 1: Assess the feasibility of merging data within each year. The purpose of this step was to consolidate the number of datafiles within a dataset for a particular year. This creates a simpler data structure and greater technical ease when performing higher-level merges (i.e., merging multiple years of data for a dataset, merging data across datasets). The team determined it was valid and appropriate to merge datafiles within each year for all datasets, using geographic identifier (GeoID) as the key merging variable.
- Step 2: Assess the feasibility of merging data across years. This was a straightforward process, given all datasets included in the project are annual data collections. The team determined it was valid and appropriate to merge datafiles across years for all datasets, using GeoID and Year as the key merging variables.
- Step 3: Assess the feasibility of merging data across datasets. This process highlighted the role of technical and conceptual linkages for cross-dataset analysis. The team determined it was valid

and appropriate to merge data across datasets using GeoID and Year as the key merging variables, for side-by-side analysis. For example, a state administrator may want to pull all nutrition services data for all OAA programs in their state to track program performance or report to relevant stakeholders. In this case, the AGID data user utilizes technical linkages to pull data from across datasets and review or analyze these data together.

The New Editions/IMPAQ team also determined it was valid and appropriate to merge data across datasets, using GeoID and Year and Variable Name, to directly compare select pairs of variables across datasets that are conceptually linked. For example, the state administrator may want to compare the OAA Title III program and OAA Title VI program in terms of the number of home-delivered meals served per participant. In this case, the AGID data user utilizes conceptual linkages to directly compare data from across datasets.

The ACL DR I project provided the foundation for aligning measures to permit cross-dataset analysis. Our efforts to advance aligning measures work during the ACL DR II project are described below.

Aligning Measures Reviews - ACL DR II

Most of the datasets included in the ACL DR I project were also included in the ACL DR II project, and the underlying data collection tools for these datasets did not change. However, the annual program performance report forms for SPR, Title VI, and NORS did change. The purpose of the aligning measures task under ACL DR II was twofold:

- to review and verify ACL's crosswalk documentation of variable changes from legacy to current report forms for SPR, Title VI, and NORS, and
- to perform aligning measures reviews of SPR, Title VI, and NORS current report forms with each other and with the other datasets included on the ACL DR II project. Conceptual linkages information for pairs of datasets that used the same data collection tools in the ACL DR I and II projects were not re-reviewed.

ACL Redesign of OAA Datasets from Legacy to Current Data Collections

Following the ACL DR I project and prior to beginning the ACL DR II project, ACL completed a redesign of the annual program performance report form for the SPR, Title VI, and NORS datasets. This redesign marked a shift from the so-called legacy data to the current data, and a transition from separate data collection and information management systems for each of these datasets to a single system, called OAAPS. The redesign involved reviewing the legacy report form for each dataset and choosing to keep, revise, or add items. ACL also made one significant methodological change. Previously, OAA Title VII grantees reported aggregated data for cases and complaints handled by Ombudsman programs. Under the redesigned NORS data collection, Ombudsman programs provide specific data for each case and complaint. These data can then be aggregated to the state level.

During the redesign, ACL aligned measures across these three datasets, where possible, and developed extensive documentation with detailed definitions so grantees can measure and report data as intended by ACL, and consistently over time. These steps have significantly improved conceptual linkages across datasets.

For example, the legacy SPR report form measured client race based on mutually exclusive race categories that captured unduplicated counts (i.e., White alone, Black or African American alone, Asian

alone, American Indian or Alaskan Native alone, Native Hawaiian or Other Pacific Islander alone, other race alone, multiple races). This approach had some important limitations. If an AGID data user wanted to know how many Black or African American clients received certain OAA Title III services, the "Black or African American alone" total would likely be an underestimate because clients who identified as Black or African American and other race(s) would be counted in the "multiple races" category. The "multiple races" category did not allow an AGID data user to determine the specific combination of races that are represented among clients counted in this category.

Additionally, an AGID data user may want to compare an OAA Title III services population to the general population of older adults in their state to determine if the services are reaching the right older adults. For example, an AGID data user might want to compare the percentage of older adults who identify as Black or African American in the OAA Title III home-delivered meals population versus the general population of older adults in their state. However, the ACS PUMS data collection follows a different measurement approach to race. These ACS PUMS data capture race as a duplicated count (e.g., Black or African American alone OR in combination with other race or races). Given the difference in measurement approach for SPR and ACS PUMS, it was not valid and appropriate to conceptually link these data using the legacy SPR data.

To address this shortcoming, ACL revised the race measure in the OAAPS to capture duplicated person count. This change addressed the limitation in the legacy SPR report form and made it valid and appropriate to conceptually link race information with ACS PUMS.

Review of Legacy to Current Data Collections and Crosswalk Documentation

During the redesign of the annual program performance report form for OAA datasets, ACL carefully documented differences between the legacy and current data collections using variable crosswalks. These crosswalk files are listed and described for SPR, Title VI, and NORS in the appendices for each dataset in Section 1: Data Documentation.

As part of the ACL DR II project, the team conducted separate reviews of the legacy and current report forms to identify linkages over time between the legacy and current versions of the data for each OAA dataset and confirmed our findings against ACL crosswalk files. Overall, the crosswalk files were comprehensive and accurate. No errors or issues were found.

One notable finding from the review was that some linkages are still technically possible between the legacy and current versions of the data despite there not being a one-to-one variable correspondence between the two versions. These linkages require additional data processing steps. For example, the legacy SPR report form collects age information on service recipients using the following age categories: 60–74, 75–84, and 85 and above. The current SPR report collects more detailed age information using the following age categories: below 60, 60–64, 65–74, 75–84, 85 and above, and age missing. Although there is only a one-to-one correspondence for two age categories in the legacy and current versions (i.e., 75–84, 85 and above), one can construct counts for service recipients ages 60–74 in the current data by summing the counts for service recipients ages 60–64 and 65–74.

Findings from our review of legacy to current data collections and crosswalk documentation for OAA datasets are provided in the file called ACL_DR_II_Crosswalk_Review.xlsx. This workbook includes separate worksheets for SPR, Title VI, and NORS. Each spreadsheet lists variable changes from the legacy to current versions of the data, notes about linkages, and key considerations from the team and ACL.

Aligning Measures Spreadsheets

The New Editions/IMPAQ team reviewed pairs of datasets to determine valid and appropriate conceptual linkages. Throughout this process, reporting forms, technical documentation, and dataset codebooks from AGID and OAAPS were consulted. Once a conceptual linkage was identified, relevant variable information was captured in an aligning measures spreadsheet, with each variable classified by topic area. Through an iterative approach, draft aligning measures spreadsheets were submitted to the ACL and the proposed conceptual linkages were discussed during project meetings. ACL also provided these spreadsheets for review and feedback to relevant ACL program staff members beyond the core ACL team members who worked on the ACL DR II project.

The New Editions/IMPAQ team then integrated all feedback into these spreadsheets and documented any rationale for determining that two variables were or were not conceptually linked across datasets. All aligning measures work is represented in the ACL_DR_II_Aligning_Measures_Spreadsheets.xlsx workbook. This workbook contains a total of 14 aligning measures spreadsheets, one spreadsheet for each pair of datasets that contained any conceptual linkages. The following pairs of datasets did not contain any conceptual linkages: Title VI–ACS PUMS, NORS–CIL, NORS–ILS, NORS–NSOAAP, NORS–ACS PUMS, NSOAAP–CIL, and NSOAAP–ILS. Exhibit 9 below lists the dataset pairs and the topic areas of variables that were conceptually linked between the two datasets.

Exhibit 9: Aligning Measures Spreadsheets

Dataset Pair	Topic Areas With Conceptual Linkages
SPR–Title VI	Staffing; Nutrition Services Persons Served; Supportive Services Persons Served; Caregiver Services Persons Served; Nutrition Services Units; Supportive Services Units; Caregiver Services Units
SPR–NORS	Total Expenditures
SPR–NSOAAP	Nutrition Services Persons Served; Supportive Services Persons Served; Caregiver Services Persons Served; Demographics; Living Alone; Functional Limitations
SPR–ACS PUMS	Demographics; Living Alone
SPR–CIL	Total Cases/Clients Served; Demographics
SPR–ILS	Total Cases/Clients Served; Demographics
Title VI–NORS	Total Expenditures
Title VI–NSOAAP	Nutrition Services Persons Served; Supportive Services Persons Served; Supportive Services Units
*NSOAAP–ACS PUMS	Demographics; Living Alone; Functional Limitations
*ACS PUMS–CIL	Demographics
*ACS PUMS–ILS	Demographics
*CIL–ILS	Demographics; Total Cases/Client Served; Service Type; Increased Independence; Improved Access; Cases Closed; Funding; Persons with Disabilities Served

*Indicates that the data collection tool for both datasets in the pair did not change since completing the ACL DR I project. Conceptual linkages for these pairs of datasets were carried over from ACL DR I project final deliverables.

Aligning Measures Summary of Findings

Conceptual linkages tended to fall under common topic areas. These topic areas provide an organizing framework to think about how measures tend to align across datasets. Exhibit 10 below lists and describes these topic areas.

Exhibit 10: Topic Areas for Conceptual Linkages

Common Topic Area	Description
Staffing	Number of full time or part time staff
Funding	Funds received from different sources, including federal and local sources
Expenditures	Total program funds expended and service-specific expenditures
Aging and Caregiver Services	Numbers of persons served and service units delivered for nutrition services, supportive services, and caregiver services
Cases, Complaints, and Requests	Numbers of cases, complaints, and requests for services received and closed by service type
Demographics	Characteristics of individuals, such as age, gender, race/ethnicity, etc.
Functional Limitations	Number of individuals served by count of functional limitations as measured by difficulties with performing ADLs and IADLs
Disability	Number of individuals served by disability type or limitation
Performance Measures	Measures of program performance, such as increased independence and improved access to services

These common topic areas are further discussed below, with examples of conceptually linked variables across aging, disability, and Census datasets included in the ACL DR II project.

Staffing

Administrative datasets often collect information on the number of full-time and part-time staff. For example, staffing information for SPR and Title VI can be conceptually linked because both datasets measure the number of full-time and part-time staff using matching definitions. More specifically, both datasets specify that the counts should include paid staff members and define 35 hours per week as the minimum for full-time employment.

Although NORS also collects staffing data, other datasets do not conceptually link to NORS because it captures combined state and local level staffing, while SPR captures staffing separately by State Unit on Aging (SUA) and Area Agency on Aging (AAA). Furthermore, these measures are not comparable because some duplication will occur across datasets as some SUAs operate Long-Term Care Ombudsman Programs (LTCOPs) directly. Also, some AAAs operate LTCOPs, and their staffing will likely be reported in both places. Additionally, not all local programs are housed at AAAs.

Funding and Expenditures

Funding information is no longer captured and reported for SPR, Title VI, and NORS under OAAPS. Only expenditures information is available from these datasets. However, CIL and ILS do capture funding information from a variety of federal government, other government, and private funding sources that is conceptually linked. There is one notable exception: the CIL dataset measures specific types of private funding sources that are not present in the ILS dataset (i.e., foundations, corporations, or trust grants; donations from individuals; membership fees; and investment income/endowment).

Expenditure terms differ for SPR, Title VI, and NORS. However, after consulting with the ACL staff, we clarified that total service expenditure, grand total dollars expended, and all funds expended can be conceptually linked. This conceptual linkage should include a note to AGID data users that total expenditures aren't measured in exactly the same way across these three datasets. The linkage is a best proxy. One additional variable in SPR and NORS appears to measure specific expenditures from OAA Title VII, Chapter 3 funds. However, it is up to a state to determine how to use their OAA Title VII, Chapter 3 funds. Some states may dedicate the funds to support their LTCOP, and as such, would report this expenditure under NORS. Other states may use the funds to support their adult protective services activities or provide education programs to older adults, and they would then report the expenditures in SPR. Due to the latitude states have to direct these funds and report them as expenditures in either SPR or NORS, the measure cannot be conceptually linked.

Aging and Caregiver Services

A variety of aging and caregiver service variables for persons served and service units are conceptually linked between SPR, Title VI, and NSOAAP data. The aging service categories persons served and service units that conceptually link between SPR and Title VI include congregate meals, home-delivered meals, case management, homemaker, personal care/home health aide, chore, and nutrition counseling. However, some important differences apply. For SPR and Title VI, the following aging services variables link only on service units: nutrition education and information/assistance.

When comparing SPR and Title VI with NSOAAP data, the total persons served data are conceptually linked for the following service categories: congregate meals, home delivered meals, case management, homemaker, assisted transportation, and caregiver. The service units are conceptually linked for assisted transportation. Notably, there is an important nuance in the way transportation service units information is measured across datasets that affects conceptual linkages. Both SPR and Title VI report service units for transportation in terms of one-way trips, but SPR reports service units separately for transportation and assisted transportation. Title VI does not differentiate between assisted and unassisted transportation. In order to validly conceptually link SPR and Title VI on transportation service units, the data for assisted transportation service units and regular transportation service units in SPR must first be summed, or 'rolled-up', before it can be validly conceptually linked with Title VI transportation service units.

Functional Limitations

Conceptual linkages exist between SPR and NSOAAP. SPR captures the number of zero, one, two, and three or more functional limitations among persons served based on the Katz Index of Independence in Activities of Daily Living (ADL)[5] and the Lawton Instrumental Activities of Daily Living (IADL) Scale.[6]

[5] The Katz Index of Independence in ADL, commonly referred to as the Katz ADL, is an assessment tool designed to evaluate functional status as a measurement of an older adult's ability to independently perform six activities of daily living (i.e., bathing, dressing, toileting, transferring, continence, and feeding). Source: Best Practices in Nursing Care to Older Adults, The Hartford Institute for Geriatric Nursing, New York University, College of Nursing. Available from: https://www.alz.org/careplanning/downloads/katz-adl.pdf

[6] The Lawton IADL Scale is an assessment tool designed to evaluate an older adult's independent living skills. These skills are considered more complex than the basic ADLs measured by the Katz ADL. There are eight domains of function measured with the Lawton IADL Scale (i.e., telephone, shopping, food preparation, housekeeping, laundry, mode of transportation, medication management, and ability to manage finances). Source: Best Practices

NSOAAP can also be used to estimate functional limitations among persons served to conceptually link with SPR. ACS PUMS captures information on disability but not on functional limitations.

Demographics

Demographics is one of the main topic areas of overlap between aging data from SPR and NSOAAP, Census data from ACS PUMS, and disability data from CIL and ILS. NSOAAP collects the respondent's date of birth to construct an age variable. A categorical variable can be generated by the user to match the age categories found in SPR and the disability datasets CIL and ILS. ACS PUMS collects age and date of birth data the same way as NSOAAP. Gender and race/ethnicity are conceptually linked across all five of these datasets.

Between ACS PUMS and NSOAAP, the following data elements are also conceptually linked: education, military status, household status (living alone, living with spouse, living with non-relatives, living with children, living with other relatives), marital status, and household income. For some elements, the specific data collection categories in ACS PUMS are more detailed than NSOAAP. For example, when asking about whether a person has ever served on active duty, ACS PUMS has additional response options for current active duty, past active duty, or active duty for the Reserves or National Guard. NSOAAP only has binary yes or no response options. The AGID data user may combine these detailed response options in ACS PUMS to create a conceptual linkage with NSOAAP.

Living alone is also conceptually linked for ACS PUMS and SPR, as well as NSOAAP and SPR. Both ACS PUMS and NSOAAP collect detailed household information that includes the total number of people who live in the home and their relationship to the older adult (e.g., relative, type of relative, non-relative). However, SPR only measures whether the older adult lives alone or not. Therefore, SPR is conceptually linked to ACS PUMS and NSOAAP only for living alone information. ACS PUMS and NSOAAP are conceptually linked for living alone and relationship information.

Other linkages

The disability datasets CIL and ILS have additional conceptual linkages related to the cases served, cases closed, and service type, along with performance measures like increased independence and improved access to services. For service type and increased independence, there are many variables that match by type, but individual elements have some slight differences in wording. Variables also measure the number of consumers overcoming barriers to their independence by accessing previously unavailable transportation, health care services, and assistive technology through independent living programs.

At first glance, the number of cases in NORS appears to link to the cases reported in the disability datasets. However, since the comparison between the NORS cases/complaints and the CIL and ILS consumer service records does not show clearly matching definitions, no conceptual linkage is established.

Key Considerations and Caveats for Conceptual Linkages

During aligning measures reviews, the New Editions/IMPAQ team found that some variables were measured similarly but were not comparable enough for conceptual linkage. For example, it appears SPR and ACS PUMS collect similar data regarding urban/rural populations. However, ACS PUMS does not

---

in Nursing Care to Older Adults, The Hartford Institute for Geriatric Nursing, New York University, College of Nursing. Available from: https://www.alz.org/careplanning/downloads/lawton-iadl.pdf

directly collect information about where the respondent lives, only the public microdata area (PUMA) code of the respondent. These PUMAs may include both rural and urban areas, and so we cannot determine urban/rural information for each respondent.

In addition, some datasets may collect the same variables but use different units in their reporting. An example is staffing, which is conceptually linked between some but not all data sources. Staffing is reported in NORS by combined count of state and local level full-time equivalent staff members, while SPR reports staffing counts by SUA and AAA separately.

Also, datasets like NSOAAP and SPR both capture information on OAA Title III services, but SPR is an administrative data collection and NSOAAP is a complex survey. NSOAAP requires weighting to estimate results that are representative at certain geographic levels. The sampling design for NSOAAP actually uses counts from SPR as a benchmark. The relationship between these datasets raises a key consideration about which source to use when a variable is measured in both datasets. Through discussions with ACL, it was determined that each estimate has value and utility to different AGID data users. Some may want to use the administrative estimate collected via the census approach used by SPR, and others may want the weighted estimated collected via the complex survey approach used by NSOAAP. Both estimates should be available to the AGID data user with a description of this consideration so that they may choose the estimate that best suits the needs of their specific analyses.

Strategies for Improving Conceptual Linkages

Advancing conceptual linkages between datasets housed on AGID should be a continuous process of reviewing and revising these data collections to improve opportunities for cross-dataset analysis and the overall utility of AGID for a variety of data users. Several strategies for improving conceptual linkages for cross-dataset analysis are provided below.

Leverage Existing Approaches for Conceptual Linkages

As AGID scales to include new datasets, ACL can leverage processes for reviewing and improving conceptual linkages on the ACL DR I project, ACL DR II project, and redesign of OAA grantee annual program performance report forms. During the ACL DR I and ACL DR II projects, the team conducted systematic reviews of pairs of datasets to determine the availability of valid and appropriate technical and conceptual linkages. A data structure for cross-dataset analysis using these linkages was designed and verified through testing. This data structure is designed to be flexible and scale to include additional datasets that may be added to AGID in the future.

The New Editions/IMPAQ team approach can be used as a procedural model for onboarding additional datasets. As new datasets are identified and onboarded into AGID, they should undergo a standard process that includes a review to determine the availability of actual linkages across datasets as well as the presence of potential linkages that could become actual linkages with revision to the measures that are potentially linked. Integrating this step into the onboarding procedure will help to maximize the number of conceptual linkages available to AGID data users at the time new datasets are added. It will also help with tracking potential linkages so that these measures can be furthered aligned.

Additionally, ACL conducted coordinated dataset reviews for SPR, Title VI, and NORS to redesign these three report forms. Through coordination, ACL reviewed and revised these data collections and significantly improved the number of conceptual linkages available among OAA datasets and between

OAA datasets and other datasets on AGID. ACL also developed detailed and comprehensive crosswalk documentation and variable definitions.

These resources help to justify the validity and appropriateness of conceptual linkages for two key reasons. First, the documentation provides clear guidance to grantees on what exactly to measure and report in their annual data submissions to ensure the data are truly aligned across datasets. Secondly, the documentation provides clear information to AGID data users about what is being measured across datasets so they can make confident determinations about whether these measures are sufficiently aligned for the purposes of their specific cross-dataset analyses.

Consider Federal Standards and Guidance for Aligning Measures on Common Topic Areas

Aligning measures across datasets will require collaboration by stakeholders responsible for each dataset. In some cases, aligning measures across datasets will not be possible given the unique information needs of the dataset and other factors, such as specific measurement requirements defined by the authorizing legislation for the data collection. However, there are many common topic areas across datasets where measures can be defined consistently to allow for conceptual linkage. Stakeholders should consider available federal guidance or standards for measuring such topic areas. For example, in 2011 HHS established data collection standards for five demographic categories: race, ethnicity, sex, primary language, and disability status.[7] These are very common topic areas measured across ACL datasets and many external datasets that may be relevant to ACL.

Following federal standards and guidance to align measures offers several benefits. These measures typically go through extensive development and testing before publication, they are widely available to the public for use in new and existing data collections, and the federal endorsement provides a standard that ACL stakeholders can follow. Additionally, these measures are commonly used, which creates opportunities for linkages with many other federally sponsored datasets as well as datasets that are not federally sponsored. More linkages increase the overall utility of the data within AGID.

Use Periodic Opportunities Such as OMB Approval Renewals to Align Measures

All federally sponsored data collections of 10 or more people require approval from OMB. The OMB review and approval process is multistaged and can take approximately 6 to 10 months to complete. Data collections are typically approved by OMB for a 3-year period. Following that 3-year period, the data collection will expire unless it is renewed, with or without changes. Members of the public are invited to review the data collection instrument(s) and provide comments.

All datasets currently on AGID are approved by OMB through this process, and it is reasonable to expect that all new datasets that may be added to AGID in the future will be required to have OMB approval. This periodic review and approval process provides a natural, recurring opportunity for stakeholders to revise measures in a dataset to form new conceptual linkages for cross-dataset analysis. ACL stakeholders should use findings from their aligning measures reviews and their tracking of potential linkages to propose revisions to their data collection instruments when renewing OMB approval.

[7] The Centers for Disease Control and Prevention (2020). Disability Datasets. Retrieved from https://www.cdc.gov/ncbddd/disabilityandhealth/datasets.html

Use Ongoing Mechanisms Such as the ACL Data Council to Review and Align Measures

The purpose of the ACL Data Council is to support ACL's mission by improving the coordination of ACL's data governance, including the development of improved processes and standards for defining, collecting, reviewing, certifying, analyzing, and presenting data ACL collects through its evaluation, grant reporting, and administrative performance measures. The Council considers a wide range of issues, including a focus on data governance standards. Through this topic, the Council aims to develop standards for data collection to make ACL data more comparable across programs (e.g., collecting categories such as age, race, and ethnicity the same way; displaying not only expenditures for OAA Title III but also grant amounts) and to help ACL to aggregate select data (e.g., demographics) agencywide.

This focus of the Council is highly consistent with aligning measures across datasets for valid and appropriate cross-dataset analysis. Whereas OMB approval renewals serve as a point in time for revising data collection instruments, the Council serves as an ongoing mechanism for ACL stakeholders to consider ways to align measures across datasets. Furthermore, the Council has an interagency membership of 23 ACL staff members from eight ACL Offices and Centers with expertise in policy, programming, research and evaluation, and information systems. This diverse membership would bring many perspectives to help determine optimal approaches for aligning measures across datasets.

Conclusions

Linkages will serve as the key for conducting valid cross-dataset analysis in the AGID 2.0 system. Technical linkages exist between data that match across datasets on the basis of time and geography. Conceptual linkages exist between data that match on the basis of time and geography, but also on the basis of subject area and measurement. For the ACL DR II project, pairs of datasets were reviewed to determine valid and appropriate conceptual linkages. Throughout this process, the team consulted reporting forms, technical documentation, and dataset codebooks from AGID and OAAPS. Once a conceptual linkage was identified, relevant variable information was captured in an aligning measures spreadsheet, with each variable classified by topic area.

ACL should continue to build on its aligning measures work to identify new linkages between datasets and to revise data collection instruments so that potential linkages are fully aligned. This will be especially important as AGID scales to include new datasets. As new datasets are identified and onboarded into AGID, they should undergo a standard process that includes aligning measures review following a systematic approach similar to the one used on the ACL DR I and ACL DR II projects. For common topic areas, such as demographics, ACL should consider federal standards and guidance for how to consistently measure such topics. The ACL Data Council can be an ongoing mechanism for aligning measures work, and OMB approval renewals can serve as point in time occasions for making aligning measures revisions to data collection instrument(s).

Once AGID 2.0 is launched, ACL should continuously manage back-end data tables, features, and services in the system to ensure AGID data users have access to up-to-date conceptual linkages information. Conceptual linkages information will be relevant to many dimensions of the AGID 2.0 system. As the number of conceptual linkages changes over time, it will be important to continuously manage and update data tables, features, and services to accurately reflect the relationships between variables so that AGID data users can pull any conceptually linked data with ease for the purposes of their analyses. For example, the XML gateway service will support data calls between AGID 2.0 and other systems; it will need to be modified to account for updated conceptual linkages. Likewise, the

topical navigation classification feature will organize information about conceptually linked data across datasets into common topic areas; it too will need to be modified to reflect updated conceptual linkages across datasets.

Section 4: Foundational Requirements for AGID 2.0

Introduction and Purpose

The ACL DR II project, Foundational Requirements for AGID 2.0 task involved business analysis and requirements gathering and documentation activities to produce this foundational requirements document for AGID 2.0 – i.e., a redesign of the current AGID system. This document is accompanied by a base set of requirements to build AGID 2.0. The base set includes business requirements and user stories supported by artifacts, including process flows and proofs of concepts (POCs). These items are presented in the appendixes that accompany this document. Descriptions of these appendixes as well as next steps and recommendations are presented within this document.

ACL has invested considerable work in developing plans for of an improved, modernized AGID system. Under this task, ACL and New Editions/IMPAQ continued to support this effort by analyzing existing documentation and conducting requirements gathering sessions to create a set of business requirements. In alignment with ACL's intention to build AGID 2.0 using Agile methodology,[8] the New Editions/IMPAQ team then created a set of user stories based on the business requirements. The team also created process flows during this period to visualize concepts to aid in defining and refining the business requirements and user stories. User stories were then leveraged to create POCs. These activities and outputs culminated in this document that provides direction and artifacts for the development team that will build AGID 2.0.

The development team referenced throughout this document includes all stakeholders who will be involved in AGID 2.0 planning, analysis, building, testing, and other relevant tasks. The development team may comprise resources like the system/product owner, business analyst(s), developer(s), tester(s), and others as determined. This team should review this document and its accompanying appendices as a group and also gather input from additional relevant ACL program, technical, and security staff, and prospective AGID users to validate or further refine the requirements. Once the document is finalized, the development team can proceed with building AGID 2.0.

Approach

During this task, the New Editions/IMPAQ team collected and analyzed existing AGID-related documentation. The documentation included AGID system materials, AGID user interviews, deliverables culled from other ACL Data Restructuring tasks, AGID market research white paper, etc. The team then developed initial business requirements based on the analysis and reviewed them with ACL program staff for validation and refinement. Multiple requirements-gathering sessions were then conducted with ACL program staff and New Editions/IMPAQ program, business analysis, and technical staff to identify additional business requirements. In some instances, process flows were created to visually depict business processes in order to aid in forming and relaying business requirements. User stories were then developed based on the business requirements and reviewed with ACL program staff for validation and refinement. Last, POCs were developed based on the user stories to provide a visual sense of the basic

[8] Agile methodology involves developing an application in smaller parts, where requirements and system features evolve through the collaborative efforts of the development team. User stories are a tool used in Agile to capture a description of an application feature from an end user perspective.

structures of the AGID 2.0 features. POCs were additionally reviewed with ACL program staff to validate and refine them.

Business Requirements

The draft business requirements presented in the table in Appendix F were identified and documented throughout this task. They describe desired functions for AGID 2.0 that ACL has identified to date. In essence, the business requirements describe what ACL would like AGID to do. The table comprises the following columns:

- Prime – a unique identifier: BR (i.e., business requirement) followed by a number and a title for a primary (or overarching) business requirement for a function. The numbers are hierarchical to differentiate the primary business requirements.
- Sub – a unique identifier for a subordinate business requirement for a function. The numbers are hierarchical and indicate the level of subordinate business requirements in relation to their primary and other subordinate business requirements. Note: The Subs and Primes for primary business requirements are the same.
- Description – description of a business requirement for a function that AGID needs to possess.
- Priority – an indication of the current priority level (or importance) of the business requirement. Note: The lead business analyst for the task determined preliminary priorities based on ACL and New Editions/IMPAQ staff observation and input. These priorities should be reviewed by the development team and adjusted, if applicable.
- Comments – notes, open items, recommendations, and tasks related to the business requirements identified by ACL and/or New Editions/IMPAQ staff. In some instances, references to related business requirements are also noted.

Table 13 presents a sample of the business requirements for AGID 2.0 Data Search and Export. Data Search and Export will allow users to search the datasets available within AGID.

Table 13: Sample of AGID 2.0 Data Search and Export Business Requirements

Prime	Sub	Description	Priority	Comments
BR-1.0 Data Search and Export	BR-1.0	System must allow users to perform a data search and export of the hierarchical trees of data elements (i.e., elements) from AGID datasets across one or more years and by one or more geographic locations; and allow users to view and export resulting, desired element values in table and geographic map format, and trends in a line chart format.	High	• Open Item: Robust queries involving significant numbers of datasets, data elements, and/or geographic locations will require a high level of system processing. Limits may need to be set on the number of datasets, data elements, and geographic locations users can select at one time if system processing capacity and/or financial limitations (to purchase a system with ample processing capability) are issues.
BR-1.0 Data Search and Export	BR-1.1	System must allow users to view and select dataset(s) from all datasets available to start a data search for desired element values.	High	
BR-1.0 Data Search and Export	BR-1.2	After selecting dataset(s), system must allow users to view and select year(s) from all years available (within the selected dataset(s)) to refine the data search.	High	

User Stories

The draft user stories presented in Appendix G represent a translation of the business requirements into a functional blueprint for AGID 2.0 features. User stories describe desired features ACL has identified to date to meet AGID users' needs. Each user story is a short, simple description of a feature (or component of a feature) told from the perspective of the person who desires a capability. Matched with each user story are acceptance criteria or a checklist that indicates all that must occur for the user story to be satisfied. In essence, the user stories and acceptance criteria describe how ACL would like users to interact with AGID features.

Each user story has a unique identifier: US (i.e., user story) followed by a number and a narrative. The numbers are hierarchical to differentiate the features and their related stories. Where applicable, preconditions are indicated to note if one user story precedes another. The user stories also map to their associated business requirements. Notations are provided throughout the acceptance criteria to provide the development team with additional specifications and references (i.e., Ref.) to related user

stories, as well as to indicate those areas that need validation or further refinement (i.e., TBD, to be determined).

Below is a portion of a sample user story for AGID 2.0 Data Search and Export, which will allow users to search the datasets available within AGID.

US-1.1. As a visitor, I want to view, search, and select data criteria to conduct an AGID data search and export the results.

- Aligned Business Requirement: BR-1.0 Data Search and Export.
- I want to view a breadcrumb to let me know where I am in the system: TBD.
- I want to view a heading that reads…: TBD
- I want to view text (TBD) describing the contents of the page.
- I want to select data criteria for my data search from AGID datasets across one or more years and by one or more geographic locations.
    - I want to view all datasets available within AGID.
        - I want to view instructions (TBD) for how to select the dataset(s).
        - I want controls to select/unselect one or more datasets that I want to include in my data search. [TBD – May need to limit number of datasets a user can select, depending on system capacity/processing limitations.]
            - I want to view a list of the datasets I have selected.
    - I want to view all years that data are available for the datasets I selected.
        - I want to view instructions (TBD) for how to select the year(s).
        - I want controls to select/unselect one or more year(s) for each dataset that I want to include in my data search.
            - I want to view a list of the years I have selected.
    - I want to view all data elements in a hierarchical tree format that are relevant to the dataset(s) and year(s) that I selected. I want to view a separate tree for each dataset (i.e., data elements are grouped by the dataset they belong in).
        - I want to view instructions for how to select the data element(s).
        - I want a search control for the data elements to narrow them down based on a keyword(s) that I enter.
        - I want controls to expand or collapse each of the branches of the hierarchical tree(s) of elements. I also want controls to expand all/collapse all branches.
        - I want controls to select/unselect one or more data elements that I want to include in my data search. [TBD – May need to limit number of data elements a user can select, depending on system capability/performance limitations.]
            - I want to view a list of the data elements I have selected grouped by dataset they belong in.

Process Flows

General process flows for several AGID 2.0 features are presented in Appendix H. These flows were created to aid ACL and New Editions/IMPAQ with concept visualization during requirements gathering. They depict core—not all—functions of the features. All functions of the features are documented in detail in the user stories. The associated user story numbers are indicated for each flow.

Figures 1 and 2 present samples of the process flows for Data Search and Export, which will allow users to search the datasets available within AGID. Figure 1 depicts the data criteria selection process, and Figure 2 depicts the search results, map and chart, and export flows. The process flows are described in detail in user story US-1.1 (in Appendix G).

Figure 1: AGID 2.0 Data Search and Export Data Criteria Selection Flow

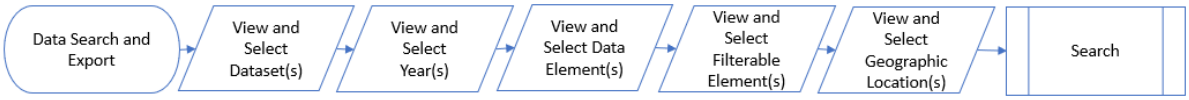
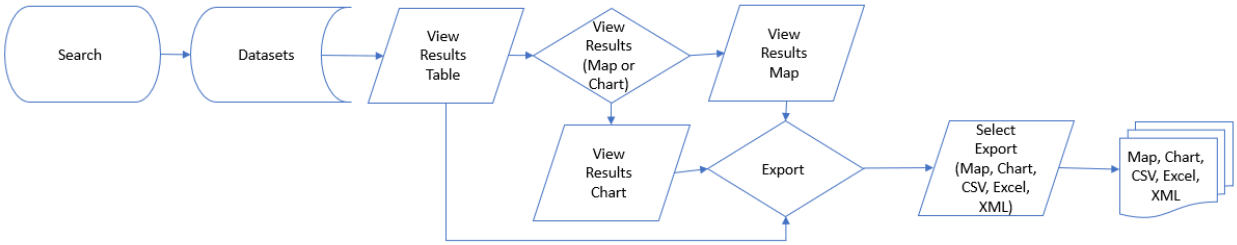


Figure 2: AGID 2.0 Data Search and Export Data Search Results, Map and Chart, and Export Flows

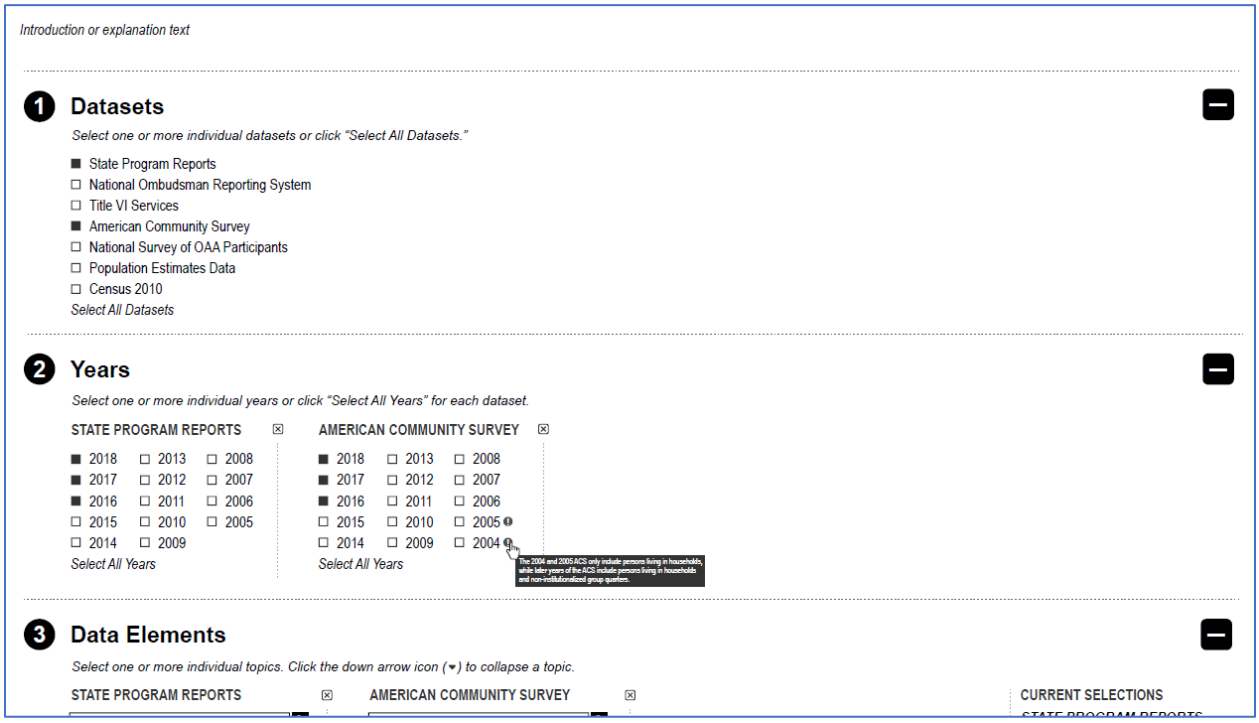


Proofs of Concepts

The POCs presented in Appendix I are diagrams of AGID 2.0 features that were derived from the user stories and acceptance criteria. The POCs illustrate the components of the AGID 2.0 features that users will experience and interact with. They are not wireframes that represent the precise placement or style of the components. Under the task, the POCs were informed by limited AGID user feedback (i.e., from AGID users who are ACL and New Editions/IMPAQ staff). In general, these users were pleased with the POCs and suggested minimal areas for improvement, and these improvements were incorporated in the POCs. The POCs act as precursors to the wireframes that the development team will create through additional stakeholder and AGID user feedback to determine placement and styling of feature components.

Figure 3 presents a cropped portion of a sample POC (i.e., Option 1 – Vertical) for AGID 2.0 Data Search and Export, which will allow users to search the datasets available within AGID.

Figure 3: Portion of the Option 1 – Vertical POC for AGID 2.0 Data Search and Export



Next Steps and Recommendations

The development team should define how to proceed moving forward with this foundational requirements document. The New Editions/IMPAQ team recommend several next steps to assist in defining an approach.

1. Address open items and tasks related to existing business requirements and identify new business requirements.

There are a number of open items and tasks in the Comments column of the business requirements table provided in Appendix F that the development team should address. Several require decisions or further analysis that may not occur prior to scoping the features and functions for the first version of AGID 2.0. In these cases, the associated functions and/or features may need to be postponed to future versions. For example, BR-16.0 Power Business Intelligence (BI) and BR-17.0 Data Lake are contingent on ACL selecting enterprise-wide platforms. Others may need to be postponed because they require further input or testing or require market research to be conducted. For example, BR-4.0 National, State, and Territory Profiles requires additional input, BR-2.0 Power Data Search and Export requires load/performance testing to determine if it is necessary, and BR-1.7.2.8 (under BR-1.0 Data Search and Export and related to mapping platforms) requires market research.

These items may need to be placed on an AGID 2.0 Roadmap as described below in 4. Reprioritize business requirements to create an AGID 2.0 Roadmap. Where indicated, the development team should create user stories and acceptance criteria for others like BR-15.0 XML Business Gateway and BR-21.0 Global/Universal. User stories for these items were not created under the task as they require input from the actual development team who will build AGID 2.0 in consultation with ACL's information technology team.

The development team should also consult with the ACL and HHS Office of the Chief Information Officer (OCIO) to determine any additional business and non-functional requirements relating to the implementation and operation of AGID 2.0. The development team should also create user stories for any of these newly identified business requirements and other technical documentation as warranted for non-functional requirements.

2. Refine existing user stories and create new user stories.

As is standard with Agile methodology, the development team should meet as a group to review the draft user stories in Appendix G and address gaps as indicated as notations in the acceptance criteria. During the review, team members may also find other areas for further refinement. ACL has also indicated an interest in testing and validating POCs based on additional stakeholder and AGID user feedback, which may result in further refinements to the user stories and acceptance criteria.

3. Test and validate proofs of concepts.

Given that ACL and New Editions/IMPAQ received user feedback from a limited number of AGID users (i.e., selected ACL and New Editions/IMPAQ staff), ACL has indicated an interest in establishing user feedback mechanisms to test and validate the POCs more broadly. The development team should consider various mechanisms (e.g., focus groups, interviews, surveys, etc.) to test the concepts with key stakeholders and AGID users. The additional feedback will help to further inform the concepts and refine the user stories and acceptance criteria. The development team may also wish to establish a user advisory group to provide ongoing feedback throughout planning, analysis, building (i.e., to review beta versions of features), and testing phases, and, potentially, periodically throughout the AGID 2.0 lifecycle to identify future enhancements. Ultimately, the development team should leverage the final POCs to create final wireframes and subsequent visuals (with the AGID 2.0 look and feel and color scheme) prior to initial programming of any features.

4. Reprioritize business requirements to create an AGID 2.0 Roadmap.

As previously noted, the lead business analyst on the task determined preliminary priorities for business requirements based on ACL and New Editions/IMPAQ staff observation and input. These priorities should be reviewed by the development team and adjusted if applicable to aid in determining which are high, medium, and low as this will aid in planning which aligned user stories (i.e., which features) have the highest priority.

ACL also may not be able to implement all business requirements and/or user stories in the first version of AGID 2.0 due to incomplete business requirements or user stories, budget or schedule constraints, etc. Therefore, the development team should create an AGID 2.0 Roadmap to aid ACL in determining short-term to long-term objectives and in planning for the overall lifecycle of AGID. New Editions/IMPAQ recommend that the roadmap be reviewed at the completion of each release version of AGID to determine if adjustments need to be made as business priorities may shift or new priorities or requirements may impact the roadmap.