



ACL's Data Quality Guidance

**Administration for Community Living
Office of Performance Evaluation**

SEPTEMBER 2020



TABLE OF CONTENTS

Overview	1
What is Quality Data?	3
Benefits of High-Quality Data	5
Improving Data Quality	6
Data Types	7
Data Components	8
Building a New Data Set	10
Implement a data collection plan	10
Set data quality standards	10
Data files	11
Documentation Files	12
Create a plan for data correction	13
Plan for widespread data integration and distribution	13
Set goals for ongoing data collection	13
Maintaining a Data Set	14
Data formatting and standardization	14
Data confidentiality and privacy	14
Data extraction	15
Preservation of data authenticity	15
Metadata to support data documentation	15
Evaluating an Existing Data Set	16
Data Dissemination	18
Data Extraction	18
Example of High-Quality Data	19
<i>Exhibit 1.</i> Older Americans Act Title III Nutrition Services:	
Number of Home-Delivered Meals	21
<i>Exhibit 2.</i> Participants in Continuing Education/Community Training	22
Summary	23
Glossary of Terms	24
Data Quality Resources	25
Citations	26

Overview

The Government Performance and Results Modernization Act of 2010, along with other legislation, requires federal agencies to report on the performance of federally funded programs and grants. The data the Administration for Community Living (ACL) collects from programs, grants, and research are used internally to make critical decisions concerning program oversight, agency initiatives, and resource allocation. Externally, ACL's data are used by the U.S. Department of Health & Human Services (HHS) headquarters to inform Congress of progress toward programmatic and organizational goals and priorities, and its impact on the public good.

The Office of Performance and Evaluation (OPE) developed this *Data Quality Guide* as an introduction to the topic of data quality and to present strategies to enhance data utility. This guide, along with *ACL's Performance Strategy*¹, aims to promote high-quality transparent information that supports sound decision-making. The *Data Quality Guide* begins by defining data quality, explaining the importance of improving data, describing data types, and providing recommendations for building, maintaining, and evaluating data sets. It then describes the role of data quality in dissemination and data extraction. It closes with examples of high-quality data and recommendations for additional resources. As you review this guide, remember, developing high-quality data is a journey and not a destination. Incremental change is worthwhile.

The Office of Management and Budget (OMB, 2001) defines “quality” as a broad term that encompasses the notions of “utility,” “objectivity,” and “integrity.” “Utility” refers to the usefulness of the information to the intended users. “Objectivity” refers to whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner and, as a matter of substance, is accurate, reliable,

¹ <https://acl.gov/sites/default/files/programs/2018-07/OPE%20PM%20Strategy%20FINAL%206-1-2018.docx>

and unbiased. “Integrity” refers to security or the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification. The General Accounting Office (GAO, 2000) adds to this definition by stating that quality data has validity, in that it adequately represents actual performance, and that it is consistent (i.e., reliable), in that it is collected repeatedly, using the same processes over time, allowing for comparisons between findings.

OPE is also available to support your office with assessing the quality of your data.

What is Quality Data?

Data collection is the process of gathering and measuring information on variables of interest in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes (Faculty Development and Instructional Center at Northern Illinois University, 2005). Data quality refers to the utility of a data set and its ability to be easily processed and analyzed.

Data must be consistent and unambiguous to be of high quality. Data quality issues are often the result of data set mergers or data integration processes in which data fields are incompatible due to schema or format inconsistencies. Data that are not of high quality can undergo data cleansing to raise their quality (Informatica, 2020).

High-quality data is consistent, unambiguous, and accurate.

The Office of Management and Budget (OMB, 2001) defines “quality” as a broad term that encompasses the notions of “utility,” “objectivity,” and “integrity.” “Utility” refers to the usefulness of the information to the intended users. “Objectivity” refers to whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner and, as a matter of substance, is accurate, reliable, and unbiased. “Integrity” refers to security or the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.²

² Office of Management and Budget. (2001). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Retrieved from https://obamawhitehouse.archives.gov/omb/fedreg_final_information_quality_guidelines/

High-quality data can be easily processed and analyzed with minimal explanations or data cleansing. High-quality data are essential to business intelligence efforts and data analytics, as well as operational efficiency. Once the data are collected, the numeric data can be used to establish performance measures that monitor the number of inputs, outputs, and outcomes (See *ACL's Performance Strategy*³). Performance measures monitor progress toward predetermined performance goals and standards, enabling program managers to make evidence-based decisions that can improve or sustain progress toward programmatic goals.

3 <https://acl.gov/sites/default/files/programs/2018-07/OPE%20PM%20Strategy%20FINAL%206-1-2018.docx>

Benefits of High-Quality Data

There are several benefits of high-quality data. First among them is the ability to communicate the relationship between inputs, action, outputs, and impact. Accurate data can “tell a story” by demonstrating to stakeholders measurable consequences of action or inaction. This information leads to informed decision-making and future action.



High-quality data lead to better decision-making across an organization. ACL’s data impact both policy and programmatic decisions. If we provide quality data, ACL leadership will have a higher level of confidence when making decisions guiding the future of ACL.



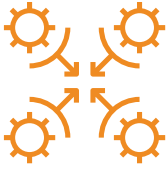
Data document the delivery of federal services and can be used to gauge effectiveness.



High-quality data enable the research community to answer research questions that have an impact on the community. Furthermore, ACL, other federal organizations, and state and local agencies depend on quality data for their policy and program planning. These policies and programs, in turn, impact many additional stakeholder groups. They, along with various advocacy groups, depend on ACL’s data to pursue funding for services and initiatives necessary to serve their constituencies.



High-quality data lead to increased consistency and coordination between existing and new data programs or grants, the elimination of redundant data (see the Paperwork Reduction Act), and the capability to analyze and visualize data.



High-quality data sets may be linked together to explore innovative and complex research topics and inform decision-making. Uniformity in design, data definitions, and identity protections make this possible.



Data advance progress on shared or similar objectives while fulfilling broader federal information needs.

Improving Data Quality

It is important to develop a systematic, standardized data collection strategy that includes a defined process for building, maintaining, and evaluating new data sets (see below). When developing or implementing an approach to improving data, give careful consideration to the needs and capacity of your program or grant. Consider addressing data challenges. Perform an in-depth analysis of existing legislation to ensure all mandatory data elements are collected, the appropriate categories are represented, and the results can be easily interpreted.

Data Types

To fulfill ACL's mission, robust, accurate information is required to inform policy and practice. There are two types of data, quantitative and qualitative. Each form of data has its advantages and disadvantages. Quantitative data provide information about quantities, i.e., numbers. Qualitative data are descriptive and describe observed phenomena, such as language. These data are nonnumerical and collected through observations, interviews, and focus groups (McLeod, 2019). Most government data, particularly performance measure data, are quantitative.

Quantitative data are typically collected using replicable, standardized processes that allow for large numbers of participants to be studied. These data are used to derive generalizations about the phenomenon being examined and, since the collection processes are standardized, it is possible to make comparisons across respondent categories over time. This approach allows for a large volume of data to be collected, but the breadth of the responses is narrower. A limitation of quantitative data is that they do not provide a rich narrative of a phenomenon. Quantitative data do not capture individual feelings and perceptions in the way that qualitative techniques can. Most population-level data collection by ACL uses quantitative data.

Qualitative data provide more depth and detail by speaking with or observing individuals rather than focusing on large groups. This approach allows for nuanced responses that, when compared to quantitative data, are more difficult to generalize and systematically compare. Due to the time required to build rapport with a respondent, a less structured data collection process, and the need to code and interpret the data, collecting qualitative data can be more time consuming (and expensive) than collecting quantitative data. Qualitative data can be transformed into quantitative data through coding procedures, though they will remain more subjective than quantitative data.

Data Components

Several key components are necessary when collecting, reporting, or assessing data, each of which can individually or cumulatively impact data quality. Understanding these components and their definitions will help to structure and evaluate the quality of data.

Validity refers to how the data are collected rather than the data themselves.

Data validation is conducted to determine which data meet a defined format, type, and range. Data are valid if they are

in the right format and of the correct type and fall within the proper range.

These characteristics help in organization and analysis. Some software automatically converts data to the correct format. For example, if you

are collecting data about the time of day users visit your site, you must decide on the format you will use. You might choose to use 24-hour time and use two digits for minutes and two for hours. Examples of this data format would include 14:34, 17:05, and 08:42. Data that do not follow this format would be invalid.

Data are valid if they are in the right format and of the correct type and fall within the proper range.

Key Components When Collecting, Reporting, or Assessing Quantitative Data

Data Dictionary	A data dictionary, as defined in the IBM Dictionary of Computing (IBM, 1993), is a “centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format.” Data dictionaries assist data users in understanding data set content and the meaning and relationships between variables.
Data Format	Formatting is an important data quality consideration due to its influence on access, preservation, and interoperability. Selecting a standard format for data allows users to access findings with off-the-shelf software, such as Excel, rather than requiring file manipulation that may introduce unwanted data errors or artifacts. The use of a standard format reduces the likelihood of data becoming unreadable as time and computer software advance. Standard formatting also facilitates research across data sets, allowing for new questions to be explored by ACL staff and external researchers. Some types of basic formatting that can improve data quality are formatting concerning currency, percentages, decimals, and dates.
Data Rules	Data rules are specific, documented standards for data entry, explaining what quality data are. An example of data quality is the use of a marital status code requiring entry of a value for single, married, widowed, or divorced. Blank entries would be considered invalid.
Percentages vs. Raw Numbers	Percentages are commonly used to describe mathematical findings. Mathematically, a percentage represents a numerator divided by a denominator (i.e., the total). A raw number is a count that has not been statistically manipulated. Raw numbers sometimes provide more transparent information than percentages. For example, you could say that 5% of a group declined an intervention. Alternatively, you could say that one out of the 20 participants declined an intervention. Both may be accurate statements, but highlighting the small number of refusals (numerator) and the small size of the group (denominator) improves the clarity of the data and the implications for the findings.
Missing Data	Missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest.
Data Analysis	Data analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.
Variance	Variance measures how far a data set is spread out. It is mathematically defined as the average of the squared differences from the mean (Statistics How To, 2020). Small variances indicate few outliers (i.e., unusually high or low numbers) in the data. Outliers can indicate errors in data collection or documentation.

Building a New Data Set

Implement a data collection plan

Using a logic model and performance measurement strategies is essential to developing a data collection plan. These practices require you to identify inputs, outputs, and outcomes; clarify definitions for each; and strategize how to collect required data and measure progress. Your data collection plan should identify the information you need to collect and additional useful information. This approach may require the establishment of additional indicators, beyond those gathered at the individual project or grantee level, and expand into the development of data-sharing agreements. Data encryption and transmission processes receive and store sensitive data, particularly data that may include personally identifiable information (PII). Explicitly state your requirements and procedures to minimize confusion and ensure timely and effective data collection. When building your data set, be mindful of the Paperwork Reduction Act (PRA) of 1995, which gives the OMB authority over the collection of information by federal agencies. The PRA is a law governing how federal agencies collect information from the public and aims to reduce the burden of collecting unnecessary data (U.S. Office of Performance Management, 2011).

It is essential to set goals for your data collection prior to developing a data system.

Set data quality standards

It is important to set data quality standards to assist in identifying erroneous data. This step should include clear instructions for the format of data entry, such as percentiles versus raw numbers, and explanations for any formulas or calculations necessary to produce data required for collection. Data quality standards should also consider outliers that may indicate data entry errors. Historical precedent,

analysis of variance, and general programmatic knowledge should inform the identification of such errors.

Data quality standards are used by ACL to certify data submissions and develop advisories for ACL's data presentation website regarding possible data quality issues. As ACL accepts data through the various grantee systems, are there clear guidelines, and are all of the grant/contract officers using the same standards for looking at the data and accepting them as complete?

One approach to establishing a data quality standard is to develop a data quality checklist. Data quality checklists can be used by a grantee to format data and reduce quality concerns for recipients and users. They can also be used by recipients, such as ACL, to identify formatting errors that may introduce incorrect or erroneous data into a data set, limiting their usefulness and accuracy. The example below, from the Oak Ridge National Laboratory Distributed Active Archive Center (n.d.), presents vital components that can be used by ACL.

Example: Data Checklist

Data files

1. Check for integrity of files (Checksum, file size, number of files)
2. Filenames are descriptive and consistent
Action: Rename data files, if needed
3. Check if file format is appropriate and can be opened
Action: Modify to archive format (non-proprietary) if needed
4. File organization is consistent and appropriate
5. Table header information complete and consistent with documentation
6. Properly versioned, if needed

Documentation Files

Documentation files describe and clarify the contents of a data file. When developing a documentation file, you should ensure that the following criteria are met.

1. Documentation matches files received
2. Data set and its contents are clearly described
3. Geospatial and temporal information are complete and described
4. Variables and units follow standards or are well defined
5. Publication or manuscript describing the data is provided
6. Methodology, calibrations, and algorithms provided
7. Known issues/limitations clearly described
8. Statements are properly referenced

The U.S. Census Bureau, as an example, provides code lists of acceptable variable responses and subject definitions for population, housing, and group quarters for when using the American Community Survey (ACS) for multiple years. It also provides guidance for comparing ACS data to 2000 and 2010 Census and instructions for applying statistical testing to ACS 1-year and 5-year data. Also, the U.S. Census Bureau provides [sample design, estimation methodology, and data accuracy explanations for 1-year and multi-year datasets](#)⁴.

Example 1.

Housing Subjects Summarized to the Block Level

(Repeated by Race and Hispanic Origin)

TABLE (MATRIX) SECTION—Con.

Table number	Table contents	Data dictionary reference name	Segment	Max size
HOUSING SUBJECTS SUMMARIZED TO THE BLOCK LEVEL (REPEATED BY RACE AND HISPANIC ORIGIN)—Con.				
H17D.	TENURE BY AGE OF HOUSEHOLDER (ASIAN ALONE HOUSEHOLDER) [21]—Con.			
	Total—Con.			
	Renter occupied:	H017D0012	46	9
	Householder 15 to 24 years	H017D0013	46	9
	Householder 25 to 34 years	H017D0014	46	9
	Householder 35 to 44 years	H017D0015	46	9
	Householder 45 to 54 years	H017D0016	46	9
	Householder 55 to 59 years	H017D0017	46	9
	Householder 60 to 64 years	H017D0018	46	9
	Householder 65 to 74 years	H017D0019	46	9
	Householder 75 to 84 years	H017D0020	46	9
	Householder 85 years and over	H017D0021	46	9

⁴ <https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>

Parameter Values

Parameter values describe the expected range and form of data, and help identify potential data errors. When developing parameter values documentation, you should ensure that the following actions are taken and criteria defined.

1. Check to ensure valid range
2. Visualize (plot, map, or both)
3. Code(s) for missing values defined and used
4. Values for coded fields defined
5. Are accuracy and precision stated and reasonable?

Example 1 from the 2018 U.S. Census shows how many of these parameter values can be communicated to a data user.

Create a plan for data correction

When identified, data errors should be investigated, corrected, or removed from the data set. Procedures should include defining the person(s) responsible for correcting data and methods for correction. This ensures a consistent process, reducing the likelihood of additional errors entering the data set.

Plan for widespread data integration and distribution

ACL is developing processes to link data sets, allowing analyses across data sets to answer new questions and provide previously unavailable insights. You should design your data set with interoperability in mind, using standardized data fields, definitions, and file formatting to allow data sets to be linked, as well as accessed and transformed using most statistical software. These practices also reduce the likelihood of introducing errors due to required data transformation.

Set goals for ongoing data collection

It is vital to remember improving data quality is a process, not an event. It may not be possible to address all data quality issues at one time. The documentation of processes and corrective actions ensure errors are addressed and not repeated.

Maintaining a Data Set

Developing and maintaining a data set takes forethought and a commitment to oversight. The results of this work are accurate data that improve organizational efficiency and productivity. For these reasons, it is crucial to investigate and assess data content, structure, relationships, and quality on an ongoing basis. Standardizing your data to meet organizational needs, as well as implementing and enforcing standardized data format rules across your organization, will establish a foundation for additional data set rules concerning data confidentiality and privacy, integrity, authenticity, and documentation. Standardized formats also enhance interoperability across data sets, allowing data to be cross-referenced and additional analyses to be performed. Data set rules can then guide data cleansing, e.g., the removal of incorrect, incomplete, duplicate, and improperly formatted data.

Data formatting and standardization

Data formatting and standardization have been previously introduced as critical components of data set development, as has planning data correction strategies. Knowing when to implement these strategies requires regular, systematic checks of data quality to ensure that data entry is adhering to formatting standards and that any data outliers are identified and checked against historical precedent, analysis of variance, and general programmatic knowledge. These strategies protect the integrity of the data. Data cleansing can then occur to remove incorrect, incomplete, duplicate, and improperly formatted data. Correcting these errors manually is inefficient and costly. Data cleansing software can systematically examine and cleanse data in a manner of minutes rather than hours, allowing program staff to focus on data monitoring, profiling, and standardization.

Data confidentiality and privacy

The federal government has a responsibility to protect the confidentiality and privacy of individuals whose data it collects, stores, and uses for analysis and programmatic

planning. HHS and ACL have strict safety, security, and privacy requirements. To ensure ACL's data maintains confidentiality and privacy, consult with ACL's Director of the Office of Information Resources Management and the Chief Information Security Officer on all safety, security, and privacy requirements. If your data is managed by a third party (a contractor or grantee), provide them with all HHS and ACL requirements.

Data extraction

Processes to access and extract information from full data sets should include mechanisms to log users and their actions within the data sets (i.e., data entry, data alteration, data downloads). A previous version of the data set should be automatically archived so that an unaltered version is available if errors are introduced during the extraction. These processes not only protect the integrity of the data but also establish an information trail useful in tracking data usage and alteration over time.

Preservation of data authenticity

Data authenticity requires content being unchanged during transmission from point to point. Inaccuracies can be introduced during transmission of a file and may not necessarily be due to tampering but merely to unattributable technology failures. The use of encryption software to authenticate that the file sent is the one received is one way to check for data origin authenticity. This data encryption also protects the data from file corruption, whether intentional or unintentional.

Metadata to support data documentation

In addition to the development of a data dictionary (previously described), data set developers should consider the use of metadata to support data documentation. Metadata provide information that describes data and, at a minimum, contains the information necessary to understand and reuse data created by others. A metadata schema is a formal framework for recording and describing data. Schemas, along with data dictionaries and readme files, which contain instructions and information about file updates, often accompany large data files and data sets (Iowa State University, University Library, n.d.).

Evaluating an Existing Data Set

ARE THE DATA RELEVANT?

Relevance is the level of consistency between data content and the area of interest of the user. To maximize the use of ACL's data, those data should not only be accurate; they must be relevant to the program's or grant's overarching goals and objectives. To ensure your data meet this requirement, link the data collection to the outcomes identified in your logic model (see *ACL's Logic Model Guidance*). It is imperative that as time evolves, your data collection evolves as well. Consider a conscientious design that anticipates future uses of data.

According to the PRA, federal agencies must aim to reduce the total amount of paperwork burden the federal government imposes through the collection of data. To achieve adherence to PRA requirements, we should only collect data that are useful and necessary.

ARE THE DATA ACCURATE?

Accuracy refers to how well the data describe real-world conditions. Accurate data convey true information. If there are errors in the data, making a data-driven decision is impossible, and your data have the potential to damage ACL's reputation. Inaccurate data create clear problems, as they can cause you to come to incorrect conclusions.

ARE THE DATA TIMELY?

Timeliness refers to how recently the event the data represent occurred. Using outdated data can lead to inaccurate results and taking actions that don't reflect the current reality. Generally, data should be recorded as soon after the real-world event as possible. Data that reflect events that happened more recently are more likely to reflect the current reality. Data typically become less useful and less accurate as time goes on.



DO YOUR DATA RELATE TO THE GOALS IDENTIFIED IN YOUR LOGIC MODEL?

Identify data needs to answer key agency questions and questions tailored to the needs of your program or grant.

ARE THE DATA COLLECTED VALUED BY STAKEHOLDERS?

Assess and balance the needs of stakeholders with the requirements and intent of the program/grant and ACL's goals and objectives.



ARE THE DATA COMPARABLE?

Data among various groups (e.g., grantees, states, etc.) should be comparable. When comparing data across multiple data sets, consistency is vital. A data item should be consistent in both its content and its format. Data items should possess the same definition and format across data collection activities and data sets. If your data aren't consistent, different groups may be operating under different assumptions about what is true. Inconsistent data can lead to departments within ACL unknowingly working against one another.

ARE THE DATA COMPLETE?

To get a full picture of the needs of programs/grants ACL supports, as well as maintain open channels of communication, our data must include all up-to-date, pertinent information. If the data are complete, there are no gaps. Everything that was supposed to be collected was successfully collected. Incomplete data may make the data reported less useful. If data are incomplete, you might have trouble gathering accurate insights. For instance, if a respondent doesn't include his or her age, it will be harder to target content to people based on their age.

Data Dissemination

ACL has a responsibility to collect, analyze, and disseminate high-quality data. The collection and analysis of high-quality data are required for programmatic oversight, evaluations of effectiveness, and mandated federal reporting. Dissemination is essential to allow additional stakeholders to benefit from data collected using federal funds. State stakeholders, including the aging and disability network, local Area Agencies on Aging, Councils on Developmental Disabilities, and other social service providers, depend on high-quality ACL data to fund and provide quality services to the community. Researchers depend on ACL's data for examining local, regional, and national service usage and service effectiveness trends and findings. Advocates and other private citizens also rely on the accuracy and access to ACL's data to persuade lawmakers and other service funders to introduce and improve funding and services for older adults, people with disabilities across the lifespan, and their families and caregivers. Ultimately, we have a responsibility to share federal agency data with state, local, and tribal governments.

Data Extraction

Data extraction involves selecting distinct data set variables and then recoding them into a format useful for analysis or presentation. Data extraction and accurate, meaningful presentation are critical to the dissemination of information and the use of collected data by stakeholders.

Be sure when developing a data collection system to discuss the required format of the data upon extraction. Consider both the type (currency, percentages, time, etc.) and the format (Excel, Pdf, Word, etc.)

Example of High-Quality Data

The American Community Survey (ACS)⁵ is a standardized data collection administered by the U.S. Census Bureau since 2005. Data are collected quarterly via the Internet, mail, telephone, and in-person interviews. The ACS addresses many of the topics within the census, but its more frequent data collection (quarterly vs. every 10 years) allows for more current data and trend analysis. The questions used in the ACS are well defined and used by many federal agencies in their data collection efforts. They can also be used at the local or regional level, as well as by researchers. ACS uses the following definition of poverty:

5 <https://www.census.gov/programs-surveys/acs>

How Poverty Is Measured

Poverty status is determined by comparing annual income to a set of dollar values (called poverty thresholds) that vary by family size, number of children, and the age of the householder. If a family's before-tax money income is less than the dollar value of their threshold, then that family and every individual in it are considered to be in poverty. For people not living in families, poverty status is determined by comparing the individual's income to his or her poverty threshold.

The poverty thresholds are updated annually to account for changes in the cost

of living using the Consumer Price Index (CPI). They do not vary geographically.

The ACS is a continuous survey, and people respond throughout the year. Since income is reported for the previous 12 months, the appropriate poverty threshold for each family is determined by multiplying the base-year poverty threshold from 1982 by the average of monthly CPI values for the 12 months preceding the survey month.

For more information, see page 107 of "American Community Survey and Puerto Rico Community Survey 2018 Subject Definitions" [here](#).

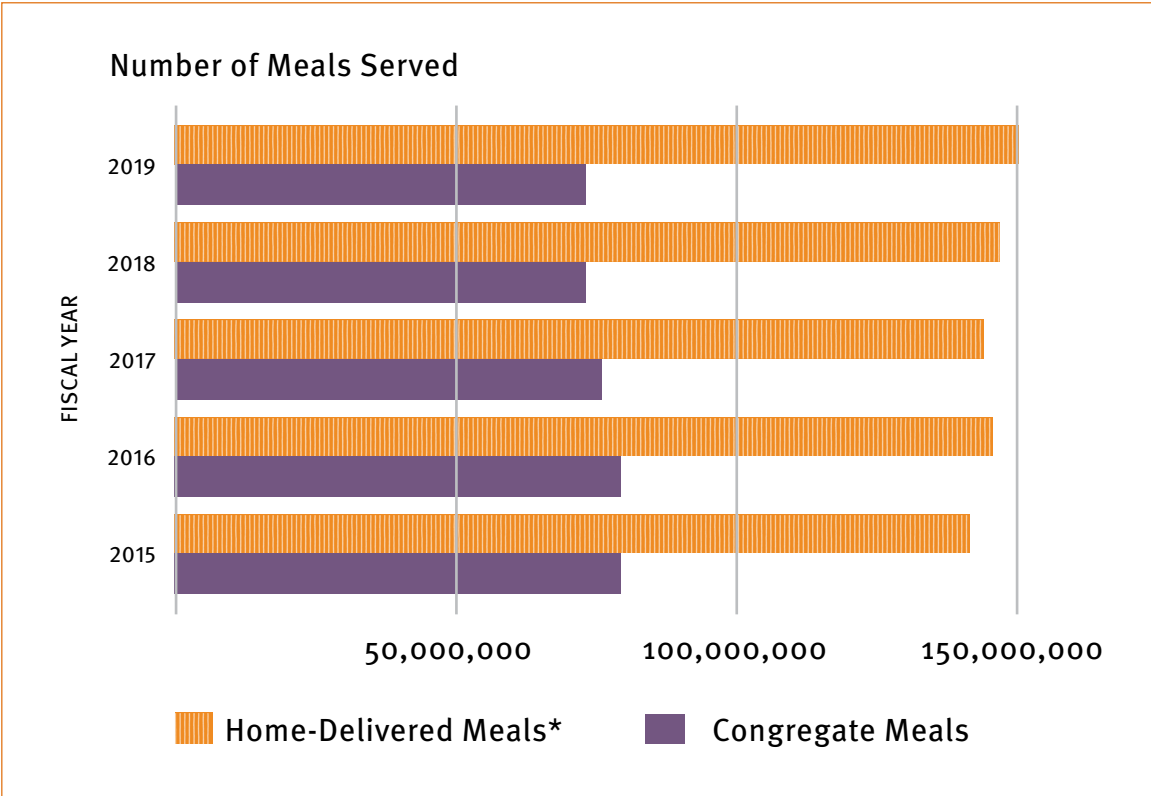
As an example of data usage, the U.S. Census Bureau examined 2017 and 2018 ACS poverty data to determine change in the national poverty rate, as well as to identify states with lower and higher poverty rates than the national average. This annual schedule allows the U.S. Census Bureau to show that there was an increase in the average poverty rate during 2007–2011, but that the rate decreased during 2012–2018. *Exhibit 1* shows the year-to-year change in the poverty rate from 2005 to 2018.

University Centers for Excellence in Developmental Disabilities (UCEDDs) are interdisciplinary education, research and public service units of a university or not-for-profit entity associated with universities. UCEDDs advise federal, state, and community policymakers and promote opportunities for individuals with developmental disabilities to exercise self-determination and to be independent, productive, integrated and included in all facets of community life. Funding from ACL provided the infrastructure support for the Centers to engage in interdisciplinary preservice training, continuing education, community services, research, and information dissemination activities. From FY 2015 to FY 2018, there was an overall increase in the number of participants receiving continued education and/or training.

ACL's Administration on Disabilities (AoD) exemplifies high-quality data through the annual use of a Tier I Review Tool to conduct high-level program compliance and outcome reviews of the Statement of Goals and Priorities (SGP) and Program Performance Report (PPR). The outcome review uses narratives and measures to evaluate performance based on data analysis to understand results. In addition to ensuring Developmental Disabilities Assistance Act compliance, AoD's Tier I Review Tool is also used to determine if a Protection and Advocacy System (P&A) needs a higher level of review such as a Tier II review.

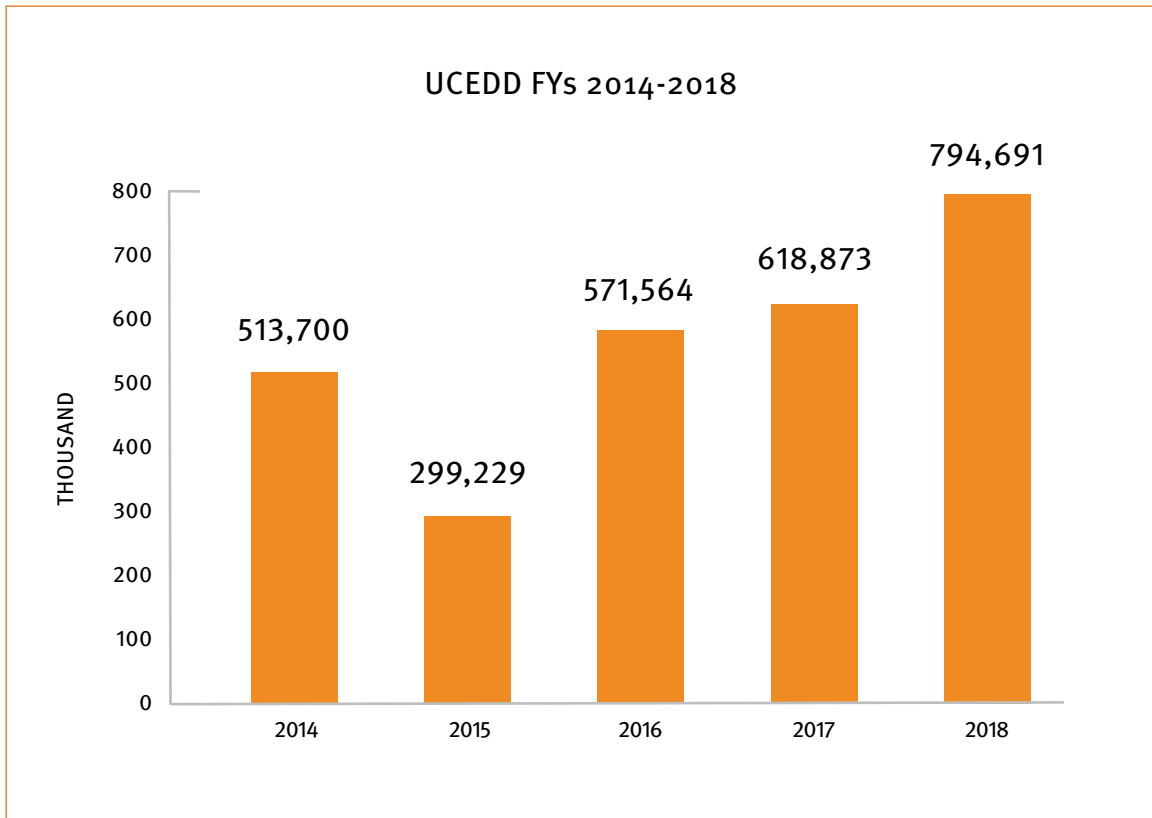
Unlike the ACS, grantees collect the Older Americans Act Title III data. The exhibit below presents information on the number of home-delivered meals distributed across the nation by federal fiscal year. Markers of quality include the clear and succinct definitions for the data, the annual collection cycle occurring in each state and territory contributing to large sample size, and the analysis and reviews occurring at multiple levels before the submission process, which enhance the accuracy of the data.

**Exhibit 1. Older Americans Act Title III Nutrition Services:
Number of Home-Delivered Meals Served by Fiscal Year**



**Home-Delivered Meal (one meal)* — A meal provided to a qualified individual in his or her place of residence. The meal is served in a program administered by State Units on Aging (SUAs) and/or Area Agencies on Aging (AAAs) and meets all of the requirements of the Older Americans Act and state/local laws. As noted in Section IIA, meals provided to individuals through means-tested programs such as Medicaid Title XIX waiver meals or other programs such as state-funded means-tested programs are excluded from the Nutrition Service Incentive Program (NSIP) meals figure in line 4a; they are included in the meal total reported on line 4 of Section IIA. Certain Title III-E funded home-delivered meals may also be included.

Exhibit 2. Participants in Continuing Education/Community Training: FYs 2014–2018



Continuing Education/ Community Training: training provided by UCEDD faculty/staff to enhance knowledge of a variety of community members (individuals with developmental and other disabilities, their families, professionals, paraprofessionals, policy makers, students or others in the community).

Summary

The collection, management, and use of high-quality data are critical to ACL and its mission to ensure older adults and people with disabilities are able to live independently and fully participate in their communities. ACL, like other federal agencies, is required to collect data to assess the performance of its federally funded programs and grants. Findings from these data collection activities are used to measure and communicate progress toward programmatic goals and priorities. Data quality is a process, and ACL is continuously seeking to improve the quality of our data through the development of formatting and interoperability standards, data transmission protocols, cleansing processes, and advisory groups such as the ACL Data Council. These evolving activities will, in time, build upon the recommendations made within this document.

Glossary of Terms

Term	Definition	Source
Data	A value or set of values representing a specific concept or concepts. Data become information when analyzed and possibly combined with other data in order to extract meaning and to provide context. The meaning of data can vary depending on its context.	The Data Reference Model: Version 2.0 https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/DRM_2_o_Final.pdf
Data Dictionary	A data dictionary is used to catalog and communicate the structure and content of data and provides meaningful descriptions for individually named data objects.	U.S. Geological Survey https://www.usgs.gov/
Data Extraction Tool	Data extraction tools allow a user to select a data basket full of variables and then recode those variables into a form that the user desires. The user can then develop customized displays of any selected data.	www.data.gov
Data Set	Datasets (or data sets) are a group of data files—usually numeric or encoded—along with the documentation files (such as a codebook, technical or methodology report, data dictionary) which explain their production or use.	Data Information Specialists Committee-UK, 2007 http://www.disc-uk.org/qanda.html
Metadata	Metadata describe a number of characteristics or attributes of data; that is, data that describe data. For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, its label, and its relationship to other data.	Adapted from The Data Reference Model: Version 2.0 https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/DRM_2_o_Final.pdf

Data Quality Resources

ACL, along with many other federal organizations, is actively engaged in improving the quality of data. While additional internal guidance is in development by ACL's Data Council, the external resources below can supplement the recommendations presented in this guide.

Federal Government Data Maturity Model

<https://www.dlt.com/blog/2018/10/08/federal-government-data-maturity-model-2019-data-management>

Federal Data Strategy 2020 Action Plan

<https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf>

H.R.150 - Grant Reporting Efficiency and Agreements Transparency Act of 2019

<https://www.congress.gov/bill/116th-congress/house-bill/150/text>

H.R.2142 - GPRA Modernization Act of 2010

<https://www.congress.gov/bill/111th-congress/house-bill/2142/text>

A Playbook in Support of the Federal Data Strategy

<https://resources.data.gov/assets/documents/fds-data-governance-playbook.pdf>

The Federal Data Strategy

<https://strategy.data.gov/overview/>

Leveraging Data as a Strategic Asset

<https://www.performance.gov/CAP/leveragingdata/>

A Guide to the Paperwork Reduction Act (PRA)

<https://pra.digital.gov/about/>

Citations

Benson, C. & Bishaw, A. (2019). Poverty: 2017 and 2018: American Community Survey briefs. Retrieved from <https://www.census.gov/content/dam/Census/library/publications/2019/acs/acsbr18-02.pdf>

Data Information Specialists Committee-UK (2007). *Datashare project, questions and answers*. Retrieved from <http://www.disc-uk.org/qanda.html>

Faculty Development and Instructional Center at Northern Illinois University. (2005). Responsible conduct in data management: Data collection. Retrieved from https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html

General Services Administration & the Office of Management and Budget. (2019). CAP goal overview: Leveraging data as a strategic asset. Retrieved from <https://www.performance.gov/CAP/leveragingdata/>

Government Accounting Office (2000). *Assessing the quality of program performance data*. Retrieved from <https://www.gao.gov/products/GAO/GGD-00-140R>

IBM. (1993). *IBM dictionary of computing*. New York, NY: McGraw-Hill, Inc. ISBN: 978-0-07 031488-7. Retrieved from <https://dl.acm.org/doi/book/10.5555/541721>

IGI Global. (n.d.). What is data relevance? Retrieved from <https://www.igi-global.com/dictionary/data-quality-assessment/6800>

Informatica. (2020). *What is data quality?* Redwood City, CA: Author. Retrieved from <https://www.informatica.com/services-and-training/glossary-of-terms/data-quality-definition.html>

Iowa State University, University Library. (n.d.). Data management plan (DMP) guide. Retrieved from <https://instr.iastate.libguides.com/dmp>

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402–406. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668100/>

Lotame. (2019). Why is data quality important? Retrieved from <https://www.lotame.com/why-is-data-quality-important/>

McLeod, S. (2019). What's the difference between qualitative and quantitative research? Retrieved from <https://www.simplypsychology.org/qualitative-quantitative.html>

Oak Ridge National Laboratory, Distributed Active Archive Center for Biochemical Dynamics. (n.d.). Data quality review checklist. Retrieved from <https://daac.ornl.gov/submit/qachecklist/>

Office of Management and Budget. (2001). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Retrieved from https://obamawhitehouse.archives.gov/omb/fedreg_final_information_quality_guidelines/

- Prell, M., Chapman, C., Adeshiyan, S., Fixler, D., Garin, T., Mirel, L., & Phipps, P. (2019). Transparent reporting for integrated data quality: Practices of seven federal statistical agencies. FCSM 19-01. Federal Committee on Statistical Methodology. Retrieved from https://nces.ed.gov/FCSM/pdf/Transparent_Reporting_FCSM_19.01.pdf
- Statistical Modeling, Causal Inference, and Social Science: Sometimes the Raw Numbers Are Better Than a Percentage. (2010). Retrieved from https://statmodeling.stat.columbia.edu/2010/06/24/sometimes_the_r/
- Statistics How To. (2020). Variance: Simple definition, step by step examples. Retrieved from <https://www.statisticshowto.com/probability-and-statistics/variance/>
- University of Virginia Library Research Data Services + Sciences. (2020). Data types & file formats. Retrieved from <https://data.library.virginia.edu/data-management/plan/format-types/>
- U.S. Census Bureau. (n.d.). American Community Survey (ACS). Retrieved from <https://www.census.gov/programs-surveys/acs>
- U.S. Geological Service (n.d.). Data management: Data dictionaries. Retrieved from <https://www.usgs.gov/products/data-and-tools/data-management/data-dictionaries>
- U.S. Office of Management and Budget (2005). The data reference model: Version 2.0. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/DRM_2_0_Final.pdf
- U.S. Office of Personnel Management. (2011). Paperwork Reduction Act (PRA) guide: Version 2.0. Retrieved from <https://www.opm.gov/about-us/open-government/digital-government-strategy/fitara/paperwork-reduction-act-guide.pdf>
- U.S. Office of Research Integrity. (n.d.). Data analysis. https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html
- Wassen, O. (2019). 11 benefits of achieving high data quality (including confidence, scalability & consistency). Retrieved from <https://www.nodegraph.se/11-benefits-of-achieving-high-data-quality/>

Developed by Jennifer Tillery, ACL Sr. Social Science Analyst
in partnership with New Editions Consulting, Inc.

For additional information contact:

Office of Performance and Evaluation
Center for Policy and Evaluation
Administration for Community Living
330 C St SW
Washington, DC 20201
www.acl.gov/contact

