Exchanging Data Using a Standards and Model Approach:

A Pilot

By Garrett County CAC and Paige Teegarden
Date: January 2018

Preface

Nonprofits must have a holistic understanding of who they serve and what services they have provided. This global data capacity spanning all programs is necessary to effectively manage their organizations and to achieve (and prove) results, which, are in turn, necessary to secure funding. Unfortunately, Community Action Agencies' (CAAs) data is invariably siloed into different databases and software systems. Data requirements are driven by different funders with different fundamental goals that require different data to evaluate results. Funders have their own data challenges and the mandate to appropriately manage their programs. In some cases, funders have required direct data entry into their own software systems. Funder mandated software (even the best designed and best intended) creates barriers for community action agencies in their efforts to better understand and serve their clients and communities. For example, CAAs can have local and privately funded programs as well as their own initiatives. Statewide software that requires direct data entry cannot be reasonably expected to meet these complex localized agency needs.

CAAs (as multi-service community-based organizations) need access to basic demographic, service, outcome and sometimes assessment data for as many of the people served as possible. This assists in making effective decisions about program/services mix and operations. CAAs must be able to uniquely identify individuals and households across programs. Further, the data models need to support local decision making on the best way to provide services in their communities.

While CAAs may be extreme in their examples, having data in a variety of siloed systems is not an uncommon problem as demonstrated by the following data standards and models, all seeking to improve interoperability:
- initiatives like NIEM (National Information Exchange Model) and NHSIA (National Human Services Interoperability Architecture),
- HL7 (electronic health standards) and
- Common Education Data Standards (CEDS) which supports the development of Longitudinal Student Data Systems
- HPXML (Weatherization data standards)

The release of the CSBG Annual Report in 2017 provided a foundational list of all data points that CAAs across the country should (if possible) collect. Beginning with this new OMB approved CSBG Report**, it is time for the Community Action Network to significantly deepen its data and interoperability capacity.**

<u>**Background:**</u>
For many years, the use of multiple client data systems, each one required by a different funder, has been a barrier to Community Action Agencies' better understanding and serving their clients and communities.

In late 2014, Paige Teegarden (VP at GCCAC) and Duane Yoder of Garret County Community Action presented a short white paper on the need for data standards in Community Action as one step toward addressing this barrier.

In April 2015, NASCSP and The Partnership hosted a meeting about data standards. At the time, the development of the CSBG Annual Report Format and ROMA Next Gen was a priority, so the work on data standards was put on hold.

The release of the CSBG Annual Report early in 2017 provided a foundation of the recommended data points necessary for CAAs to understand. At that time, Garrett County Community Action and the software vendor that they are part owners in, empowOR by CSST (headed up by Paige Teegarden), began working on the data exchange problem with a small amount of Funding from the Annie E. Casey Foundation and their own donations of time.

The following Agencies participated in the "CSBG Data Standards Pilot": Garrett County Community Action, Allegany HRDC, Douglas County, CO, Reno CAA, and Mohawk CAA. This group met for several months, facilitated by Paige, to provide input on data standards definitions. Paige drafted the preliminary model and file definitions. This group worked with CSST Software to further test the transfer of initial data from different data systems using the data standards and then aggregating data for the CSBG Annual Report. This experience was presented during the Community Action Partnership Annual Conference in August 2017. The group also kept in touch with the NASCSP Data Task Force coordinators, with the intention that the Data Standards work should coordinate with but not be a replacement for the Data Task Force work. The group also sought input from Barbara Mooney of ACRT, plus and solicited feedback at conferences.

This paper reports on the lessons and recommendations from that Pilot Project.

This document contains following sections

Pilot Process

For several months over the course of Spring 2017, members of the Data Standards Pilot provided input and assisted Paige Teegarden in completing the following:

1. Diagraming the relationships between major types of data (data entities) used to develop the CSBG Annual Report. We also included data such as assessments that might be helpful to local coordination of services and determination of outcomes
2. Took each of these 'data entities' and developed draft csv file definitions in discussion with participating agencies over the course of several months
3. Looked at HMIS, NEIM and Child Education Data Standards as a starting point for understanding common data elements and existing data models and standards and how those were created
4. Requested sample data from the participating agencies; we only asked that they provide data that would not take significant time, energy or effort for them pull from their systems
5. Using a standard BI tool (and no coding), created CSBG Annual Report aggregations from the data from different agencies and different systems.
6. Created other "bundled services" reports that looked at relationship of number of services to number of outcomes

Based on these experiences, representatives of the Pilot Group presented at the Community Action Partnership Annual Conference and included a hands-on exercise that illustrated 'why bother' with data standards.   The hands-on exercise is described in the attachments.

Developers or agency people who participated in generating the files for the pilot shared their feedback about what would make it easier to get information from systems; we also discussed other data structures that people knew about and how those structures would 'deal with' pulling data in the initial csv format. These shared experiences and subsequent discussions during the pilot were taken into consideration in the development of the proposed model, data dictionary and xsd that have been submitted for consideration as a CANDiO proposal.

Defining The Problem

Community Action Agencies' (CAAs) client level data is often stored in different software systems and databases that may not be controlled by the agency. CAAs need to understand/analyze client information may be collected in disparate programs and often stored in multiple systems.

They need this data for
   ● **reporting and analysis, and,**
   ● **for better coordination of services locally.**

These two different needs may result in different, but related, exchange standards.

Discussions with the pilot agencies revealed the following parts of the data exchange problem:

1. **"On My Own"/Duplicating Time/Need for Local Capacity**. Currently every community action agency is figuring out how to approach these issues independently and ends up having independent discussions with software vendors, funders and monitors. This requires significant local staff capacity to understand technical options and convince others to follow through. It also requires a commitment from leadership to spend time, energy and political capital in these discussions problem independently.

2. **State and Federal Funders "ownership" of data and requiring direct data entry.** Examples:
   a. Being required to enter LIHEAP data into a State controlled system that currently does not accept imports (which would simplify the local CAA's internal work flow)
   b. Being required to enter Homeless Data into an HMIS and not being able to either import or export from that system. The CAA may be in multiple Continuums.

3. **Privacy and Confidentiality Issues:** there is general confusion about what personal information local agencies are allowed to ask from people seeking different kinds of services and then what information agency staff can enter where.

4. **Data Mapping Issues** - both in terms of like concepts called different things in different systems (i.e. *Dwelling Type* versus *Housing Type*) and in terms of options (what options different funders want for the same concept or idea). We delve more into this below.

There are several technical/data aspects to the problem of exchanging data from one system to another at the local agency level which are described in more detail below.

- **Calling fields/elements by different names whether those fields/elements conceptually represent the same thing or not.** i.e. In looking at the two field/elements *HouseInfo* and *DwellingType*, it is unclear whether those fields represent the same thing or not. Another example would be *Annual Household Income*, which does not make clear whether it is gross or net.
  – By providing a specific data dictionary that explains what the field represents, with the exchange format, the responsibility for mapping one field name to the standard is placed on the vendor and should only need to be done once. Agencies remain responsible for their own data entry accuracy.

- **"List Options"/ "enumerated values" for same concept are different for different funders** – whether the options are conceptually consistent or conceptually different. An example might be an 'element' called *Housing Status* that conceptually represents the housing situation of a group of people. One group of funders might define the list of options as: *Rent, Own, Subsidized Rent, Homeless, Long Term care, Other*. Another funder might list the options as: *Permanent Rental Housing, At Risk of Homelessness, Literally Homeless, and*

Transitional Housing. It is unclear, without further guidance, how these 'fit together' or map to each other. Another example is *Family Type* where some lists of options may be: *Household with Children and Household without Children; another might be Single Parent, 2 Parent, No Children, Other*; and another may be *Single Parent Female, Single Parent Male, 2 Parent, Unrelated Adults, Related Adults, Other*.

   – Again, a Data Dictionary and enumerated list options (defined in xsd or in csv options) that can mapped to one another in a single step would provide some resolution to this issue, and would provide clearly identified areas where there are mismatches across data standards.

- **Data Relationships**. Data dictionaries can help with the first two issues, but for the meaning of data to be clear, the clarification of relationship between pieces of information is necessary. For example, benefits are related to households for CSBG reporting purposes but public benefits can also be related to individuals in reality and in data systems.

   – Consequently, the data model needs to enforce clarity about what the relationship actually is for this specific person or household in this data set. This clarity in turn allows for data analysis such as providing the x% of households receiving emergency food service who also receive SNAP benefits; and then further, of those households that do not receive SNAP benefits, what % are headed by someone over the age of 60 or what % have children under the age of 5.

   –

- **Household Composition**—a Special Case. Different Federal and State funding programs define the "Economic Unit"/Group of People who should be counted for eligibility in different ways. For example, Head Start would not count grandparents who were not contributing to the welfare of the grandchild in a household for eligibility purposes, even if the child and child's mother are living with the grandparents. But LIHEAP and other housing programs need every person living in that house for eligibility. The ability to clearly define which of these groupings is on the application/eligibility may be important. Further, since CAAs work with clients over time, clients are often in continually changing household configurations.

   – Consequently, we need not just a data model that supports clarity about the information being passed, we also need clear guidance about what information is counted for what aggregations in reports. For example, we would advocate that if a person is in a household and received a service (and perhaps an outcome), that that household is counted in "All Characteristics" (Module 4: Section C) and if that person moves into a different household and receives services that the second household is also counted.

Lessons from Pilot Experience

Following are lessons based on discussions and experiences over the course of the pilot.

- **Supportive Principles and Guidelines are needed** - Principles that support an organizational culture open to sharing information as well as Guidelines about what structures

need to be in place within an agency or collaboration to facilitate data sharing would be helpful for agencies moving to more data integration.

- **Different purposes for data sharing** - There are distinct scenarios for sharing data:
  - sharing information for purposes of deduplication, analysis and reporting
  - sharing information for coordination of services
  - sharing information within an agency or across agencies for better coordination
  - sharing data with funders and researchers

While all these scenarios face some of the same issues and could be helped by data standardization, each exchange of data should describe its purpose. In the pilot we discussed different purposes, but we only tested for the purpose of reporting.

- **Deduplication of Clients** - Deduplicating clients across an agency's programs or between multiple agencies requires a means of determining when people 'match'—i.e. when Mary Jones in one data set is the same person as Mary Jones in another data set. However, for most CAA purposes this matching (and ultimate deduplication) does not have to be perfect.
  - **Which Data Sets are Needed?** Agencies (and in some cases funders) need to agree on which sets of data (i.e. data from different agencies, data from all programs or some programs, or different data systems) are anticipated to have sufficient information provided to de-duplicate people.
  - **Additional PPI Needed** - De-duplicating clients across different data sets will require more "Personal private information" and/or human intervention than normally needed when compiling data from different data sets that each have their own unique ID for the person and household
    - Holding a discussion about social security numbers and/or last 4 of SSN with other personal information would be helpful to any exchange process.
  - **Consent / Protection of PPI** - The agencies that control the personal data collected will need to have MOUs in place to protect sharing of any identifying information; however if agencies are only sharing de-identified data, they may not need client consent.

- **Household Composition over Time** - Given CAAs focus, it is critical for systems and any data standardization (data model and standards) to support a means of identifying household composition in ways that support shifts in households over time
  - In the NHSIA data model, the approach is to record the people in the household or co-applying for human services benefits at the time of application or enrollment and not to seek systemic tracking of household composition through time.

- **Different Levels of Focus** - Stakeholders– agency front line, directors, executive directors, private funders, state monitors and federal partners—have different interests in sharing data that need to be understood at the local level to help build support and understanding of context and requirements for data exchange (see tips below for more on this).

- **Data Standards & Transfer Options must be Made Clear** - Funders, federal and state monitors and others engaged in decisions about how to gather and receive client data need to understand options for receiving data and the role that data standards and data stores can play in allowing them to receive the accurate data they need without forcing agencies to directly enter client data into a specific software system.

- **Common Language Base for Client Assessments** - Assessments used by local CAAs with their clients vary tremendously and should be flexible enough to support local differences. The dimensions in the CSBG Annual report (employment, housing, education, asset/income, health/behavior, safety, civic engagement/leadership), and the 5 categories in ROMA training (Crisis, Vulnerable, Stable, Safe and Thriving) provide a common language and base to share information across agencies for further analysis.

- **Data Model must be Flexible** - Any CSBG Data Model needs to be flexible enough to both exchange data locally for coordination of services and to provide guidance on the aggregation of data for reporting and analysis.
    - A CSBG Data Model needs to support a subset of information collected and tracked by programs commonly operated by CAAs. A CSBG Data Model could be extended for programmatic specific management and operational needs but need not include all elements of every program's operational and management need for every program. For example, agreeing on data collected by LIHEAP that can be provided in a standardized form compliant with a CSBG Data model might include applicant name and demographic information (in CSBG form) but the CSBG model itself need not standardized every element of Utility Vendor and Benefits
    - A CSBG Data Model should reference other models as possible.
    - During the Pilot, information was passed via CSV file. While that will likely continue to be one option for exchange, more sophisticated exchange methods including xsd/xml will help to validate some data and should be part of standardization.

- **Data Model must support Existing Diverse Software** - Software systems have diverse data models and relationships including a couple described below. Any CSBG data model needs to support these differences assuming they could all be accurately used for analysis and reporting:
    - Income can be model in at least the following (and probably other) ways:
        - Income sources (i.e. employment, SSI) are each recorded separately with a start and end time for each individual.
        - A "snapshot" / summary of all income on particular day for a person or household that also includes amount of income by source
        - A total household income on a date, with an indication of whether it includes income from specific sources, but without providing the amount of that source
    - Public Benefits might be related to a person, or to a Household, or both; and might be recorded for a period of time or on a particular date.
    - Assessment questions might be related to a person or to a Household or to both

- Information about disability status, housing status, employment status, military status might be tracked over time or only include current information.

Conclusions and Recommendations

CAA's siloed and inaccessible data has a long history with many contributing factors. Based on the experiences in this Pilot, we recommend a multi-prong approach to address the issue. We further suggest utilizing a **collaborative project** that we are calling CANDiO (Community Action Network Data InterOperability) Project. We'd suggest that different stakeholders (including vendors, national partners, state monitors and local agencies) participate in CANDiO, and that it be a vendor-neutral, collaborative engaged project.

Proposed Purpose Statement for CANDiO: The CANDiO (Community Action Network Data interOperability) Project ultimately seeks to reduce the costs of sharing critical data across various systems for community action agencies (and any others that can use the same data model)—for data analysis and reporting purposes, and locally for operational coordination of services.

Recommendations for Community Action Network Wide Strategies

Strategy: Technical Standardization
We strongly believe that a **necessary step** in addressing data silos is the development of:
- Data Model
- Data Dictionary
- Data Standards for exchange (xml and rules)

Without an easy, low cost technical path forward that doesn't require each agency to address issues independently, there is little incentive for agencies to work with staff or address privacy and ownership issues.

Strategy: Pilots
Pilot and share experiences with local data stores (or data warehouses) using the above standards. Encourage local agencies and communities to use the draft xsd and data store approach and to share their experiences. One example of these kinds of pilots are those being supported by the AECF through the Whole Family Initiative with the Community Action Partnership. This strategy could include opportunities for broader information sharing about the pilots.

Strategy: Education and Peer Sharing
Part of the problem is the isolation and lack of experience with data exchanges across all stakeholders. This strategy might include:
- Learning Communities
- Share MOUs
- Share legal language or opinions about asking and recording private data
- Share best practices in security policies and monitoring

Recommendations for Local Agencies

Below are a few ideas and thoughts for CAAs seeking to increase their data capacity—their ability to collect, record, analyze and use client specific information.

- Develop and monitor client privacy, consent and security policies. Adopt the philosophy that the client owns their data and the agency must be good stewards of the information.

- Do a client data system 'scan' in your organization so that you understand and document:
  - What client-related information is collected by different programs (start with a short description you can work into details later) and where the programs record and store that data
  - Understand if these systems are required, recommended or just preferred.
  - For any required systems, formally ask if there are import or data exchange options available.

- Work with your state CSBG Offices to have them be your ambassadors within state government; helping to explain the CAA issue and encouraging their counterparts to allow imports/data exchanges for State required systems.

- Develop a list of key common data elements that are collected from anyone who is considered a client of the agency, and train all staff on collecting and recording these elements.

- Understand that if you have data in multiple systems and you want to de-duplicate those clients for reporting, you will need a means of knowing that Client A (John Doe) in system A is the same person as Client B (John J Doe) in system B. The most commonly used means of this is a social security number or the last 4 digits of the social security number in combination with several other pieces of personal information such as name and DOB. Develop guidance for staff and security policies around handling SSNs.

- You must have data in order to use data. Focus on getting clear and consistent practices in place around the collection and recording of information. Identify a few metrics that your leadership team monitors.

- Create opportunities for ongoing training and support for staff that collect, record and analyze client data.

- Distribute responsibility among staff around data collection, analysis and use so that this capacity is in more than one location in the agency

- Make the following part of your Leadership:

- o Be a role model for data use in decision making
- o Communicate expectations around what data is collected and where it is entered and by whom
- o Commit to an on-going learning process and an openness to change based on new experiences and information
- o Prioritize and commit time and resources to using data
- o Directly and personally participate in discussions about data and information
- o Commit resources for technology and staff development
- o Support staff with an aptitude for data analysis

- Understand that different stakeholders need different things from data. The following chart shows some of the many wishes / requirements of the various stakeholders at different points in the process of data collection and use:

Phases of Data☐ Stakeholders ↓	Data Collection	Data Entry & Monitoring	Data Analysis	Data Sharing and Use
Clients / Participant	Don't take too much Time	Security of my data is important	New Insights	Share with me & Get consent
Frontline Staff	Relationship w/ client & their duties	# Software(s) entry Time (ease of User Interface)	Include me when doing analysis	Make staff job easier
Management	Correct, consistent info. for program management	Monitor data quality		Monitor consent, negotiate MOU
Leadership	Correct, consistent, useful info	Reduce double data entry	Support capacity Must ask about data supporting decisions	Intentional time to analyze and discuss
Funders	Program specified data	Support data to data exchange	Fund local agencies use of technology and building their data capacity	Allow innovative solutions

ATTACHMENT

Prior to meeting/training, take two boxes one should have no divisions in it; the other should be divided into 6 sections and labeled with Q1, Q2, Q3, Q4, Q5.

Divide the room into two groups.

Ask one group to Write the answer to following – each on separate card.
1. What's your Family Type
2. What's your Favorite Animal
3. How much money do you have in your wallet?
4. What service are you seeking? What do you need today?
5. List any publicly funded benefits that you currently receive.

Second Group
Take 5 Cards and Write your Initials and DOB on upper right corner of each card
Write Q# AND your answer to each of the following questions on a separate one of those cards

Q1. Select the following "Family Type" that best matches people currently in your household:
>> Single female parent
>> Single male parent
>> Two parent (or parent figures)
>> Grandparent(s) with Grandchild
>> Grandparent(s), Parent(s), Child
>> Adults no children
>> Unrelated Adults
>> Other

Q2. Select one of the following that best represents your favorite animal.
>> Dog
>> Cat
>> Bird
>> Horse
>> Other

Q3. How much money do you have in your wallet?
>> <$10
>> $ 10 to $50
>> $51 to $100
>> Over $100

Q4. Select ALL of the following that you need help with today:
>> Housing Assistance
>> Place to stay tonight
>> Energy cut off notice

» Childcare
» Want to buy a home
» Job Skills Training or Coaching
» Services for an elderly adult
» Services for a disabled adult
» Food
» Medical Services

Q5. Please select any of the following publicly funded benefits that you currently receive.

» TANF
» WIC
» SNAP/Food Stamps
» Medicaid
» Affordable Care Act Subsidy
» Medicare
» Childcare Subsidy

Group one , throw all cards into the unsegmented box. Group two, place answers to each question in appropriately segmented place in the segmented box.

Give Group 1, group 2's box; give group 2, group 1's box. I want a summary of who is in the room and what they want.

NOTE:
Data standards help with knowing what data to store where; dictionaries help with defining options.
Data Standards and Models Provide:
– "lexicon" /common language
– structure (think about the difference between cards in one box and cards with initials and DOB in box with sections
– 'options' that help define what that data is representing
– relationship between data elements (what information 'holds' other information)
– "Rules of the road" for what kind of data is expected in what format and with what options...

Resulting in faster (and therefore cheaper) and more accurate combination of data which in turn allows:
– Analysis and research
– Reporting different systems
– Do your specific work while giving people access to that information in the form they need