Introduction

Since the introduction of ROMA guidelines in 1994, and particularly since the beginning of mandated outcomes tracking in 1998, community action agencies (CAAs) have worked to balance two imperatives: delivering critical services and collecting comprehensive data about their programs' impact on clients. These tasks are complicated by the tendency of funders, administrators, and other stakeholders to require specific data systems for specific programs. Community Action Agencies nationwide use numerous differing systems to track client information. This fragmentation of client data poses considerable difficulty when organizations try to assemble an overall picture of the populations they serve, whether for reporting purposes (e.g. the annual CSBG IS Survey) or for internal performance assessments. Acquiring an "unduplicated client count" is a tough, sometimes unobtainable goal requiring considerable expense in staff time.

Background

Community Action Agencies are nearly always providers of multiple services, funded by a heterogenous set of grants. In addition to anti-poverty efforts backed by the Community Services Block Grant (CSBG[1]), CAAs commonly provide Head Start[2] and other early childhood services, low-income heating assistance and home weatherization (LIHEAP[3] and WAP[4]), nutrition programs like WIC[5] and SNAP[6], as well as youth employment, housing, elder care, and a host of other services. States administer the bulk of these programs and in many cases mandate specific software systems to assist in tracking service delivery.

For example, Rhode Island's Department of Human Services[7] manages the state's federal LIHEAP and WAP grants using specialized systems developed by Hancock Software[8]. The LIHEAP program is locally administered by seven community action agencies throughout the state, who are responsible for client outreach and the coordination of service delivery with local energy providers. Every client application and service detail is entered into Hancock's system, and the full program commanded more than $27 million in funding for the 2015 fiscal year[9]. But for the CAAs providing the service, LIHEAP is just one of the tools they use to help their clients out of poverty. Many of the

[1] CSBG, U.S. Dept. of Health and Human Services (http://www.acf.hhs.gov/programs/ocs/programs/csbg)
[2] Office of Head Start (http://www.acf.hhs.gov/programs/ohs)
[3] Low Income Home Energy Assistance Program (http://www.acf.hhs.gov/programs/ocs/programs/liheap)
[4] Weatherization Assistance Program (http://energy.gov/eere/wipo/where-apply-weatherization-assistance)
[5] Women, Infants, and Children (USDA - http://www.fns.usda.gov/wic/women-infants-and-children-wic)
[6] SNAP (USDA - http://www.fns.usda.gov/snap/supplemental-nutrition-assistance-program-snap)
[7] Rhode Island Department of Human Services (http://www.dhs.ri.gov)
[8] Hancock Software (http://www.hancocksoftware.com)
[9] LIHEAP FY 2015 Funding (http://www.liheapch.acf.hhs.gov/profiles/RI.htm)

same individuals have children enrolled in Head Start, are receiving case management services, or live in subsidized housing. As with LIHEAP, each of these programs has its own database for tracking and administering client services.

There are good reasons for software specialization when it comes to administering human services. Programs like LIHEAP and Head Start have different target clients, different eligibility rules, and different anticipated outcomes. Systems like Hancock's are fine-tuned to handle the details around home heating and serve as a conduit to energy providers, but they lack features required by other human services programs. So while a good LIHEAP database can be a wonderful resource for the state agency that oversees the program, as well as the community action agencies responsible for its day-to-day administration, it can't fulfill the needs of every program offered under the CAA umbrella.

As a result, community action agencies inevitably accumulate databases. While some programs can share a single database, many of the larger grants require the use of discrete, purpose-built systems. The challenge for the CAA is that these separate systems end up siloing information, making attempts to report on overall agency activities challenging. And this leads to a larger problem of demonstrating the effectiveness of community action programs in the struggle against poverty.

There are very real technical problems that arise when attempting to provide a unified view of data stored across multiple systems. How are duplicate client records identified and handled? What do the data fields in each system signify, and can they be made to relate to the other systems? For a multi-service agency that serves individuals and households enrolled in multiple programs throughout the year, how can a comprehensive, accurate report of these activities be developed?

Codect™

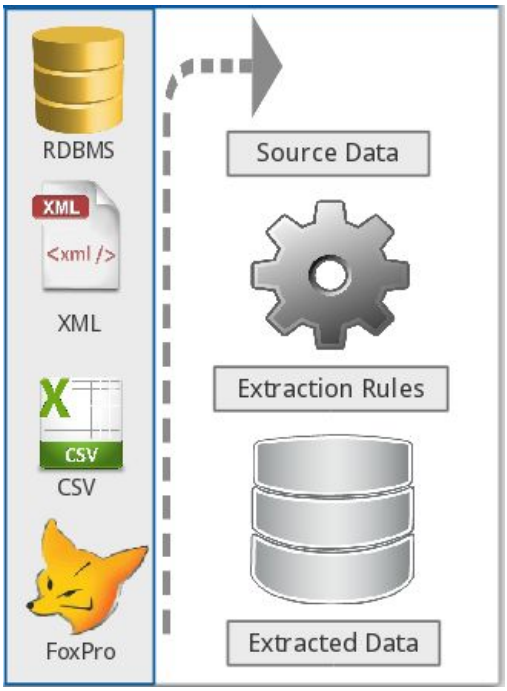
Fulcimus, LLC has developed a solution to the problem outlined above, the Codect™ data warehouse. Codect is designed to securely integrate data from any source system and provide community action agencies with a set of comprehensive reporting tools that tell the full story about the work they do. Since human services data are also highly sensitive and fall under numerous state and federal privacy rules, Codect has built-in end-to-end encryption, implemented to exceed current standards[10]. Codect offers an easy-to use web interface, as well as a library of reports, including the CSBG IS Survey.

Codect has two major components: the **Codect Connector**, a modular and easily-modifiable application that extracts data from source systems and the **Codect Data Integrator**, which transforms, cleans, and de-duplicates incoming data into a final database.

[10] Codect uses LUKS/dm-crypt AES256 to protect data at rest and 2048-bit SSL to protect data in transit.

In terms of architecture, Codect's design uses a modified version of the Extract, Transform, Load (ETL) method employed by many data warehousing applications, as well as a denormalized "big data" approach to handle the heterogeneity of extracted data. The application is entirely written in the Java programming language, and it uses the well-respected open source database PostgreSQL for storage.

Codect Connector



The Codect Connector (*outline to the right*) is the first stage of the data warehousing process and is the program that retrieves (or extracts) data from the various systems used by community action agencies. Designed for flexibility, the Codect Connector can be configured to recognize and extract data from any digital repository. As of this writing, the application recognizes XML, CSV (comma-separated values), MySQL, Microsoft SQL Server, Microsoft Access, and Microsoft Visual FoxPro legacy data sets, along with the data structures and logic of numerous human services and electronic health records systems.

While Codect is a fixed product, the complex and changing nature of the service it provides necessitates a design that allows for adaptability over the long term. Database vendors alter their products, and state agencies transition to new vendors for their programs as needs change. The Codect Connector permits new data sources to be added and existing data sources to be changed without impacting other components of the warehouse. Data extracted from source systems are flattened into a common format and prepared for later stages of the warehousing process. Once in this format, no further information about the design of the source system is required. To the best of our knowledge, this is the most flexible solution ever developed for integrating community action agency data.

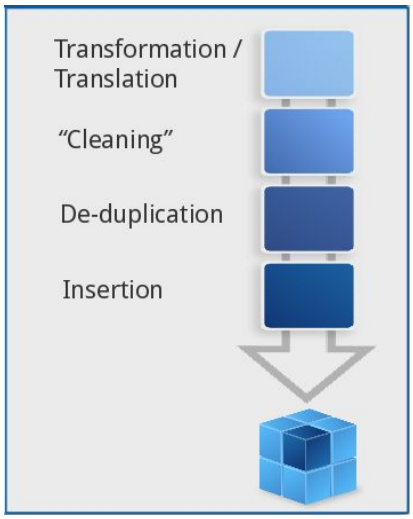
Codect Data Integrator

The Codect Data Integrator provides the "transformative" phase of the data warehousing process, where raw extracted data made available by the Codect Connector are processed. The diagram below gives an overview of the sequence of steps executed by the application.

The first step, Transformation and Translation, turns specific coded values from source systems into a common format. For instance, one system might encode gender with a 1 for female and a 2 for male, while another might use the characters 'F' and 'M' for

the same information. At the end of this step, the numerous specific codes used by each source system are translated to the Codect data warehouse's format.

The second step, Cleaning, verifies the validity of extracted data. Given the volume and varying quality of source data, obvious errors need to be identified and removed before reaching later stages of the process. Common examples include birth dates set in the future and codes that don't conform to a source system's data definition. Once this step has completed, the clean records are run through a process of De-duplication where individuals and households are identified and matched across the full set of source data. The final step is Insertion: The de-duplicated records are then loaded into the warehouse where they can be accessed for use in reports, particularly agency-wide aggregate reports like the CSBG IS Survey. Finally, Codect offers a web front end with a user-friendly dashboard that presents all of this information for review.



Beyond the Silo

Codect offers community action agencies the ability to reconcile the use of mandated or purpose-built databases for their programs with the broader requirements of agency-wide reporting. It frees agencies from committing staff time to tedious and error-prone duplicate data entry tasks in an effort to compile cross-service numbers, allowing them to focus their resources on service delivery. By bringing disparate data together, it helps them get the most out of each of their program-specific databases. With Codect, the story of community action agencies' valuable contributions can be clearly and unambiguously told.