Chapter 2

Tips and Techniques for Linking Multiple Data Systems: The Illinois Department of Human Services Consolidation Project

John Van Voorhis, David Koepke, and David Yu
University of Chicago

Abstract

This project involves the linkage of individuals across more than 20 state-run programs including TANF (AFDC), Medicaid, JOBS, Child Protection, Child Welfare Services, Alcohol and Substance Abuse programs, WIC, and mental health services. The count before linking is over 7.5 million records of individuals. Unduplicating the datasets leaves 5.9 million records. And the final linked dataset contains records for 4.1 million individuals. This study will provide the basic population counts for the State of Illinois's planning for the consolidation of these programs into a new Department of Human Services.

In the context of linking multiple systems, we have done a number of different things to make using AutoMatch easier. Some features of the process relate to standardized file and directory layouts, automatically generating match scripts, "data improvement" algorithms, and false match detection.

The first two issues, files and directories and scripts, are primarily technical, while the second two issues have more general substantive content in addition to the technical matter.

Properly laying out the tools for a matching project is a critical part of its success. Having a standard form for variable standardization, unduplication and matching provides a firm and stable foundation for linking many files together. Creating additional automation tools for working within such standards is also well worth the time it takes to make them.

With multiple sources of data it is possible to improve the data fields for individuals who are linked across multiple datasets. We will discuss both how we extract the information needed for such improvements and how we use it to improve the master list of individuals. One particular example of these improvements involves resolving the false linking of family members.