

MAINTAINING TRACKING INFORMATION ON SUBJECTS IN LONGITUDINAL STUDIES

Sherry Brown-Scoggins & Sandra Rothwell
National Center for Health Statistics Abstract

Introduction

Over the past decade, a number of surveys within the National Center for Health Statistics (NCHS, Center) have added longitudinal components. For all such studies, it is necessary to "track" survey subjects and proxy respondents for several years (up to 20 years in some cases). The purpose of tracking is to maintain current address and telephone information, verify administrative identifiers such as MEDICARE numbers, and determine accurate vital status data. A high survey follow-up rate is essential to obtain reliable study results. While all tracking activities have much in common, leaving each survey to plan and execute its own tracking effort would result in a lack of consistency from survey to survey. In addition, each such data collection effort would incur the cost of developing tracking strategies and supporting software.

Hence, the decision was made to develop a software system, the NCHS Automated Tracking System (NCHSAT), which would facilitate passive tracking and would be applicable to any survey conducted by NCHS. The system automates the clerical functions of a number of tracking activities frequently used for Center surveys. These functions include creating both paper request forms and electronic files, based on a tracking schedule; capturing, reviewing and processing the information returned from the tracking source; and updating a tracking data base.

System Design

Over the past decade, development of NCHSAT software system has shifted from SAS/AF on an IBM MVS mainframe (1990 - 1997) to client/server development via PowerBuilder/Sybase on a Unix platform (1996 - present). The shift to client/server development was driven by a need to 1) include more surveys; 2) implement a more flexible system design; 3) increase speed of data access; 4) eliminate the need for mainframe tapes; and 5) reduce errors in system use.

A large Sybase database is used to maintain information on every subject, proxy or tracking contact included in NCHS longitudinal studies. There were over 113,000 persons being tracked in the mainframe system. An additional 160,000 were added in 1997. All of this data will be migrated from SAS to Sybase by the end of 1999. Each person in the data base is identified as a participant at one or more levels of at least one, but possibly more than one, survey. This adds layers of complexity to data management. Additionally, two other issues complicate system development: 1) custom needs of individual surveys, and 2) the dependency of passive tracking on the storage, availability and maintenance of both current and historical data. The client/server database environment provides the necessary flexibility in data storage and access, and the resources required to support the expansion of passive tracking activity.

System development is phased. Each phase is designed for each source of tracking information. Sources include the National Death Index (NDI) used to update vital status and obtain cause of death; the National Change of Address Registry (NCOA) which updates addresses for individuals who fill out forwarding orders at the Post Office; and Post Office verification (POST) which updates address information in response to verification forms sent to Postmasters.

System Characteristics

- Year 2000 Compliant
- Client/server using PowerBuilder front-end and Sybase relational database on Unix server
- Supports an unlimited number of surveys
- Phased development; each tracking source becomes a system module
- Additional features include security layers, data import/export options, a variety of reports