

SOCIOMETRICS PROJECT No. 50

ESTABLISHING A CONTEXTUAL DATA ARCHIVE PHASE I

This Small Business Innovation Research (SBIR) Phase I project tested the feasibility of establishing a **Contextual Data Archive**, which would facilitate access to the best contextual data. The archive would comprise three kinds of machine-readable materials: (1) the principal data sets with the potential for use as sources of contextual data; (2) research materials—a collection of key publications and documents relating to the use of contextual data; and (3) software to find and retrieve those portions of the data sets and research materials relevant to a user's needs. The archive would offer all its products in a variety of machine-readable media, including CD-ROM.

Data Sets. The heart of the archive would be approximately 20 data sets that contain contextual data describing geographic areas from the size of the block to the state. Each data set in the collection would be archived according to Sociometrics' standardized conventions that have been designed to make the data sets as accessible and easy to use for secondary analysis as possible, thus encouraging their use. Each data set would include a machine-readable raw data file; machine-readable text files providing the set-up programs for creating SPSS and SAS system files for several operating platforms; and a machine-readable as well as printed User's Guide describing the content and format of the data.

Research Materials. The archive would also contain a collection of publications and documents on the use of contextual data, available in both machine-readable and printed formats. The collection would emphasize two kinds of materials: research reports of analyses using contextual data in large demographic data sets; and methodological papers that address the methods and limitations of contextual data analysis. Emphasis would be given to materials that were exemplary in their use of contextual data or in their treatment of key methodological issues in contextual analysis.

Software. The archive would make available search and retrieval software allowing the user to: (1) identify variables across the entire data archive meeting user-specified criteria of topic, type, and keyword, and then (2) create analysis sub-files that contain only the desired variables. The software would, in addition, allow (3) full-text keyword search of user's guides, measurement instruments, and research materials.

CD-ROM. The entire collection of data sets, research materials, and search and retrieval software would be placed on a single CD-ROM.

Phase II was subsequently funded (See Project 58).

Sponsor:	National Institute of Child Health and Human Development Demographic and Behavioral Sciences Branch Center for Population Research, 6100 Executive Blvd., Room 8B-13 Bethesda, MD 20892 (301) 496-1174	
Grant No:	1 R43 HD32290-01	Program Official: Christine Bachrach
Duration:	01/01/95—07/31/95	Budget: \$70,770
Staff	J. L. Peterson, V. Christensen, D. Kim	