Running records and the automated reconstruction of historical narrative

Beveridge, Andrew A.; Sweeting, George V.

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-34785

RUNNING RECORDS AND THE AUTOMATED RECONSTRUCTION OF HISTORICAL NARRATIVE(*)

Andrew A. Beveridge and George V. Sweeting(+)

Abstract: All social science research is concerned with the study of change, yet most scholars who use quantitative methods - historians, sociologists, and economists alike - have been unable to exploit fully data sources which illuminate how social change actually occurs. We have developed a methodology, known as "running records"(1), which allows researchers not only to use data sources, such as census schedules and tax lists which are essentially static, but also by exploiting the power of the computer, to utilize the vast of materials which record the events and transactions of everyday life for large populations - sources which document the process of change itself, as it relates to existing structures and institutions.

This paper reports on the use of our "running records" methodology, which has wide application. The procedure takes as its main focus the actions of groups and individuals. In essence it uses the power of the computer to analyze narrative records in ways analogous to the practices of "traditional" historians but for sample sizes and populations familiar to "new" social historians and quantitative social scientists. This methodology allows one to use a computer to organize and analyze materials as a conventional historian might. The operations that historians and social scientists perform - scanning through documents, looking for commonalities, assigning them to categories - are assisted by the computer.(2) We will describe this methodology, briefly compare our approach with some other historical data base systems, and give some examples of its uses.

METHODOLOGICAL APPROACHES OF SOCIAL SCIENTISTS AND HISTORIANS

Sociologists and historians using quantitative methods have seldom attempted to base their analyses on action or narrative. Instead, they use a variety of approaches each of which mainly uses materials that are static.(3) The most common approach, of course, is cross-sectional. Data is collected at one date, usually from one source, using one instrument. Historical or contemporary censuses are good examples, and survey analysis is the classic example. At its best, such a method can give a fairly accurate picture of the state of a people or society at one historical point. For example, one can know the relation between income, occupation, education and father's occupation at some date in the past. The processes underlying that relationship cannot be adequately studied using only cross-sectional material.(4) Nonetheless, researchers have attempted to make inferences from such static data sources. In a very basic sense, the substantive need to be able to understand social structure and social mobility has outpaced the methodology needed to analyze it. Similarly, one might know a great deal about the social structure at one time, for instance the distribution of wealth in the

(+) Address all communications to: Andrew A. Beveridge, Queens College of the City University of New York, Department of Sociology, Flushing, New York 11367

colonies on the eve of the Revolution. How such a distribution came about and evolved could not be examined directly.(5) In short the possible explanations for the distribution cannot be adequately assessed using data from one point in time. Instead, various hypothetical processes can be proposed.

Some scholars have attempted comparative cross-sectional analysis, but one still faces the same basic problem. Comparison allows us to see what changes have occurred in a society, but they can not explain why or how the change took place. For instance, many different processes could significantly alter the proportion of members of a given ethnic group who held one particular occupation in the 1850's.(6) Yet none of these processes can be assumed from comparing two cross-sections. Similarly, the fact that father's occupation became less associated with son's occupation over a ten year span, gives only limited insight into why this is so.(7) Still in all, comparative statics or comparative cross-sectional analysis does begin to make the process of change central to the analysis. The state of a community, society or organization is depicted at two or more points in time. Thus, the issue that becomes important is manner in which the system changed from the first to later points. What processes caused it? Methods for attempting to infer change from such types of information have been developed by economists and other social scientists.(8)

Time series analysis is a complex application of comparative cross-sectional or comparative static analysis. Many different measurements are made of the state of a system of many different points in time. Although many data points exist, conventionally one is looking at the joint change in aggregate variables. Any of these could be subject to a variety of processes. For instance, a particularly high cotton price could be the result of a draught or disruption in world cotton markets.(9) Rapid migration to non-metropolitan areas in the U.S. could be due to a variety of factors. The forced integration of certain schools might make some individuals move. However, one needs other information to truly test such a hypothesis. Yet, time series analysis because of its many different measurement points often does make it possible to rule out certain types of relationships, which might seem to be correct for two or three comparative cross-sections. In effect, various statistical techniques are used to attempt to analyze the processes underlying change, because it cannot be measured directly.(10)

Social scientists and others have recognized problems with these common approaches. Data collection efforts have been mounted which follow given units (usually, individuals) during long periods of time. In some cases, attempts to reconstruct the complete work or marital history of individuals have been made. Such efforts have led to interesting, even surprising, findings about the stability of individuals' life situations and how changes among individuals are related to general changes in the social system.(11) Sophisticated new tools have been developed to model such changes. New findings have been gleaned about such topics as the impact of economic status on marital behavior, the determinants of unemployment, and the spread of municipal reform in the United States. The series of events that an individual or organization might undergo is related to past events, as well as with other individuals and organizations using "events history" analysis, developed by Nancy Tuma and Michael Hannan.(12)

Even methods which take into account the "events history" do not reflect completely the structure of the relationships in which the events are embodied. Events can be interpreted in a variety of ways depending on the surrounding relationships. To put it simply, historical change is very

complex. Individual actions and transactions, as well as the way in which such actions and transactions are structured all can be important. Because of this complexity, existing methods and existing data sources are sometimes confounded with a l l of historical reality. For example, if a scholar has collected information from several censuses and a few business records, he or she will have a natural tendency to conflate such data along with the analytical techniques he or she is using with important historical change during a given period, even if this means ignoring most actions, interactions, and direct family and other relations. The incomplete picture will be confused with the whole portrait. If the scholar's approach is adopted by other researchers in other settings, other pictures will be painted in very similar hues for other communities as the same painting materials, palette, brushes and easel and style is used by others.

This is not to say that scholars necessarily ignore or misconstrue the process of change. Indeed, many make compelling arguments about the character of the underlying process. Still in all, the areas where safe generalizations can be made are often defined by the sources in existence. In relying so heavily on survey schedules, tax lists, census manuscripts, probate records and the like, scholars have by-passed sources which record transactions and interactions and social and institutional relationships, as they occur. We call these sources "running records". We maintain that analyzing such materials organized in an accessible form (even without recourse to complex statistical modelling procedures) has and may lead to very different interpretations of the course of historical change, than are common today. Indeed, we would completely agree with the note that the Macfarlane group published in Historical Methods a number of years ago:
> "Our preliminary findings suggest that the whole framework which has been developed during the last 40 years by sociologist and historians to explain the transition to the "modern" world is both crude, and in many respects, mistaken ... Without the very detailed analysis of many thousands of interactions and transactions shown by overlapping documents, it would have been impossible to realize how very inapplicable are the general sociological models of transition in this case."(13)

In short, running records fill in the gaps in accounts of historical and social change not by recourse to complex statistical models based on small amounts of data, not by abstruse uses of apriori theoretical models, but by a general method of recording and analyzing varieties of historical materials. The "running records" methodology also has within it the possibility of verification or disconfirmation. Findings generated by one researcher can be examined by others without necessarily carrying out further archival research. New interpretations may be tested by using existing data bases, or by merely adding a small fraction of additional materials to them.

WHAT ARE RUNNING RECORDS?

Running records record actions and transactions by and between people and institutions. They usually are generated as part of the normal day to day operation of some organization or bureaucracy. For instance (to steal an example from the project that spawned these methods), if a bank gives a loan to an individual the entry in the bank's ledger book recording the loan (and usually including other information about the loan) constitutes a single running record. Sampling from a single page, or even a few pages, of the ledger book (analogous to cross-sectional data gathering) would tell little about the role of the bank in the community because only a few transactions would be recorded on a given page. However, a researcher utilizing the entire ledger book (or a significant part of it) would be able to begin to explore important questions concerning access to credit, the relationships

between loan recipients and the bank, and, when combined with data from other sources, the importance of credit in facilitating economic development. Similarly, to understand the changing role of a textile or steel mill in a community requires more than information about the firm, its owners, and its employees at one point in time.

An individual's social status or category often changes as a result of the transactions noted in a running record source and thus such change can be specifically linked to a given action. For instance, one might receive a loan or a government contract in part because of personal or other ties. Such goods can lead to wealth or high social status. This change in status can be traced to a decision made by others acting out their social roles under the influence of personal and other social relationships. As another example, workers might recommend kin to their employers, and their employers might well prefer to hire the kin of their workers. Such actions, which seem to us to be very commonplace, could easily leave some firms with a distinctly "ethnic" work force. Once again, this is the case of a social category or state resulting from action. In other words, running records are "narrative" history, whereas sources such as census manuscripts offer only a "snap shot" of the community at a single point in time without showing the web of relationships which produced the image caught in the picture.

Tools developed to analyze such cross-sectional materials by their very nature not take into account the "running" nature of time which pervades all human and social relationships. It is for just such situations that running records hold the key to more powerful and meaningful analyses. We do not claim that the use of the types of sources we call "running records" is a significant breakthrough per se; they have always been used by some scholars. However quantitative historians and social scientists have largely overlooked their potential because of the difficulty of organizing them for large samples and populations. We believe that our methodology overcomes this problem. The essence of using running records is the concept that a record runs, or is valid until it is over-ridden by a later event. For instance, an individual is alive until he or she dies, is married until he or she becomes widowed or divorced. Similarly, one may hold one job until he or she takes another one and so on. A person may also be in several different states at the same time. The problem the researcher must overcome is to be able to infer the state at any given time from the single record of a change in status. We have done this using computerized techniques which draw upon both hierarchical and relational data base models. We will describe our approach in greater detail below.

There are many types of running records which could be utilized by quantitative historians and social scientists to study actions over time. Some examples include:
1. Personnel records - which follow careers over time;
2. Business records - which record actions and transactions of one or more businesses;
3. Genealogies - which are action records noting marriages, births and deaths for families;
4. Court and public records - where individuals become involved in one or more activities;
5. School records - where early careers are recorded;
6. Some government and voluntary association records - where status is conferred;
7. Land transactions - where property changes hands repeatedly.

In all cases "running records" provide information on actions and transactions among many individuals and on the relations among people and institutions.

IMPORTANCE OF MULTIPLE SOURCES

Our running rcords methodology exploits new types of materials by using different data handling techniques. It also relies upon multiple record sources. Using many sources is admittedly not a major innovation from the perspective of the traditional narrative historian who is trained to seek out as many different perspectives on a historical event as archival holdings permit. However, much quantitative historical and social research has been based on a few (in some cases single) record sources; sources which often suffer from the limitations described above. Single source studies are based upon assumptions about the completeness and accuracy of any one record source - assumptions which the experience of generations of historians has shown to be optimistic at best. Because the survival of historical records is so selective and uncontrollable and the quality of those which survive is often untestable in isolation we believe that it is incumbent on researchers to gather data from as many sources as possible - even quantitative historians and social scientists working with large samples or populations. The programs we have developed use the power of the computer to reduce somewhat the extra time and effort required to meet this obligation, while allowing opportunities for systematic analyses of questions which otherwise might seem intractable.

Macfarlane offers an excellent description of the virtues of multiple sources.(14) Each source contributes small pieces of information about an individual although the value or meaning of that information may not be apparent at the time. Slowly, through an additive process, the available data about an individual is built up from many sources. The new information illuminates the old and vice versa, often in ways that could not have been foreseen when the first information was recorded. As Macfarlane points out, "time and man's negligence have selectively destroyed very large parts of the past; other parts were never committed to paper. What remains is infinitely precious and each tiny shred needs to be carefully used."(15)

The use of multiple sources goes a long way, as well towards overcoming the perennial problem of nominal record linkage. Although any one source may have biases or errors which exclude some individuals, by bringing to bear the information available from several sources the chances of correctly identifying people and entities are greatly increased. It may well be that part of the remarkably high geographic mobility rates reported for nineteenth century American communities may really be accounted for by the inability of researchers relying primarily on manuscript census schedules to find people ten years later because of limitations of the census as a record source.(16) Using the data collected for Beveridge's study of Cheshire County, New Hampshire (described below) we have compared the persistence of small businesses with the persistence rates of similar firms reported for Poughkeepsie, New York.(17) We found that persistence in Cheshire County measured using data files which exploited our multiple sources was significantly higher than Hutchinson, relying on a single source, reported for Poughkeepsie. Furthermore, when we simulated the use of a single source for our data base the peristence rates fell to approximately those reported by Hutchinson.(18) Clearly, multiple sources if nothing else, increase the chances of tracking individuals in the past over a longer period of time.

Yet multiple sources, particularly if they include some of the types of running records described above, do more than help researchers identify people in the past. They also give quantitative historians and social scientists the chance to analyze communities or societies from a perspective which reflects the complex and multiple relationships which make up social structures, a perspective much more accurate than that afforded by a few static census schedules or tax lists. As Macfarlane points out, "We are interested in human beings and their activities in as many contexts as possible ... The essence of the approach is the necessity that several different records bear on a particular individual at different points in his or her life."(19) It is only when these records are put together and logically integrated by quantitative historians or social scientists that we can really understand the processes of social change.

Macfarlane's efforts at community reconstruction, although based on a sound methodological approach, have been only partially successful at this writing. He has candidly admitted that the task proved too large for their non-computerized data system.(20) They began the laborious process of computerizing their materials, both by indexing the original documents and by entering transcriptions of those documents, into the computer.(21) To adequately use running records requires computerized data base methods which 1. use multiple sources (some of which should run), 2. code the full text of a piece of data for easy retrieval and recoding, 3. preserve a record of network relationships which can be recreated for analysis, and 4. allow simple modeling of change over time. Such an effort permits the type of reconstruction of the past which Macfarlane, Hershberg, and others have attempted. Our methodology, described below, seeks to meet these objectives.

SOME ALTERNATIVES

The complicated organization of the data in Beveridge's study of Cheshire County (see below) and the inability of existing computer packages to manage the task led to the development of our "running records" methodology. In this section we will briefly describe the limitations of some alternative approaches and then describe our system.

Packages such as SPSS, BMDP, DATATEXT, and SAS (although SAS and more recently SPSS, have now significantly enhanced their ability to handle hierarchical data structures) assume that data will be arranged with every variable in a specific column. Such a structure is known as a "flat" file. Since running records can document as many transactions as actually take place for any given individual such a "flat" structure is inappropriate. These packages also provide few tools for recording network relationships or for restructuring the data based on such relationships. Because of these shortcomings the "system" file routines of such packages could not be used for the basic data handling steps in Beveridge's study. These packages can be utilized, however, for quantitative and statistical analyses after the data has been organized by other means. Furthermore, as SAS has been enhanced over the years, more and more of our methodology has been reprogrammed using SAS. One remaining shortcoming of SAS is its inability to handle network relationships in a relatively simple manner.

In addition to packages developed specifically for social science research there are many commercial systems for data base management. Most, however, are optimized for objectives that are not important for most quantitative historical or social scientific research projects. For example, many feature a capacity to do complex retrievals of small amounts of data from a data base, while most social scientists want to analyze data from a large

fraction of the data base. Secondly many commercial systems are set up for interactive use with very large on-line data bases. Aside from convenience most researchers have little reason to use such on-line access, especially since it entails very high real costs. Finally, most commercial systems do not have facilities to handle analysis over long spans of time, yet such data and its analysis should be at the heart of quantitative historical research.

Other scholars attempting to conduct large scale social history projects also have developed sets of computer programs to solve many of the same problems we faced. Some of these are quite sound and solve some of the problems involved, especially those of nominal record linkage. Yet, few scholars have used historical running records, so it appears that none can readily handle the types of data we are using.

The Philadelphia Social History Project (PSHP), under the direction of Theodore Hershberg, developed a set of computer routines for handling nineteenth century data. These programs facilitate the entry, verification, and manipulation of historical records tracing over 500,000 individuals in Philadelphia between 1850 and 1880. However, the types of sources used by Hershberg and his co-workers are essentially cross-sectional, consisting primarily of manuscript census schedules and city directories rather than what we would consider running records. Therefore, although record linkage was a primary concern to the project, and a sophisticated linkage procedure was devised, the use of such static sources limited the number of individuals successfully traced over time. Accurately linking individuals across a ten year gulf is difficult, particularly if only the few variables recorded on the census schedules are available for matching. We believe that data sources recording events throughout the period under study, rather than cross-sectional observations every ten years, would have significantly increased the number of individuals linked. This would have made it possible to explore more fully the questions of how and why the changes in various social and economic spheres described by Hershberg occurred.(22)

The computer programs developed in connection with the PSHP are useful for handling cross-sectional historical data for very large samples. It is possible to trace and aggregate information from the records of individuals who share some relationship (i.e. reside in the same household) and to build dictionaries for easy recoding of variables. Unfortunately, these programs require that much of the structure be specified at the beginning of the project, leaving little room for incorporating new conceptualizations of the relationships between individuals which become apparent after the data collection process has begun. Furthermore, they offer little aid to researchers attempting to exploit running records because of the assumption that there will be only a few explicity defined 'observations' for any individual in the files. The PSHP programs are not appropriate for the type of research we are conducting.

Manfred Thaller's CLIO system is an ambitious attempt to create a standard set of programs which can be used by researchers studying different historical problems.(23) CLIO provides easy to use data entry, retrieval, and recoding and data transformation routines to facilitate the development of historical data bases. Yet CLIO does not provide some of the features which we feel are essential for historical quantitative social science research.

One aspect of CLIO which we view as a shortcoming others may see as a virtue - its comprehensiveness. In order for the set of routine to work together as

an overall system important decisions about the data structure and the types of relationships between individual cases (e.g. people, firms, families) are imposed by the system rather than determined by the researcher. In the interest of ease of use and generalizability, CLIO sacrifices essential flexibility, especially given the great variety of historical questions and data sources amenable to quantitative analysis. This does not mean that each researcher must be prepared or capable of designing and programming his or her own data base system. Rather, tools should be developed which combine ease of use while permitting easy definition and modification of data structures and relationships by the researcher, long after data entry has begun if necessary.

The second major problem with CLIO is that it seems designed largely to carry out the type of ahistorical research described above. There appears to be little possibility of incorporating action as an element in analysis using CLIO data files. Although the data entry routines could easily be used for coding 'running records' the data retrieval routines are oriented towards generating 'flat' SPSS type records for analysis. Once again the questions which may be answered seem limited to descriptions of a community at one or more given points in time rather than the much more interesting 'narrative' ones of how those structures came to be and how they changed over time.

OUR APPROACH TO USING "RUNNING RECORDS"

Faced with the shortcomings of existing alternatives it was necessary to develop a new approach. As a starting point a system designed under the direction of Richard Ruggles and Orin Hansen and put in use by the National Bureau of Economic Research was adopted. It was significantly enhanced and new data checking routines were added. In addition, two extremely important programs were written to allow network relationships to be recorded and then utilized for analysis.(24) Although these programs were written in connection with one research project, because the specific variables and network structures all are table driven (and hence modifiable), they all are generalizable for use with many types of historical data.(25)

The running records master files are stored as hierarchical sequential files.(26) Each piece of information is entered as a character literal and then stamped with a unique identifier. This tag includes the name of the individual or entity the data pertains to, the date of the piece of data, the type of data included, and sequential sorting information. Because of the necessity for restructuring information for various sorts of analyses, we have found it necessary to use such a long sort key field, even at some sacrifice of compactness. The use of high capacity storage media and the relatively small cost of additional sequential processing of the data files makes this a small sacrifice. In addition, the tag stamps make the printed file easily human readable which is essential both when matching existing data files with new sources, and for simple editing and recoding of the files. Because the data is entered as original character literals, variables may be recoded over and over again depending on the particular analysis being conducted.

The ability to trace and analyze network relationships, which is one of virtues of the running records methodology, draws upon the relational data model. When a record of a relationship is encountered in a data source, for example the fact that one person is the business associate of another, that fact is entered for the first individual. Subsequently a program known as XREF is run which generates a record in the second individual's subfile also

noting the relationship. Thus the relationship is recorded and traced in the master file. To use this relationship for analysis a second program, SCHEMAT, was developed which changes the identifier in the tag field of some or all of the records on one side of the relationship. Thus, after resorting the file the subfile of one of the business associates now includes some information from the subfile of the other business associate. The following example gives some indication of the potential power of this approach.

The officers of banks are often prohibited by state law or charter from loaning to themselves; however there may not be any limits on loans to their associates. If the bank officers are known and their business associates are identified it is possible to move information pertaining to loans received by business associates of officers into the subfiles of the officers themselves. This makes it possible to measure the amount of indirect self-dealing by bank officers. Because the number and nature of the relationships to be traced are open-ended and can be specified, or respecified, by the researcher at will, these routines could be used for analyzing and tracing almost any type of social network.

SOME EXAMPLES

Our running records methodology has already been successfully used in several different research settings.(27) We will describe two of them below which will further demonstrate how the approach is used.

Credit Allocation and Community Change

The project which began our interest in using running records is a study of credit allocation and change in one community over eighty years. The plainly ambitious goal of the study is to analyze the dynamics of social change by using varied sources of information. By relating actions in one sphere to those in another, the study is addressing a series of important issues. This is an attempt to study a community by systematically following many individuals from one area of life to another, for instance: between home and work and business, through kin relations, to other institutional realms, eg., church, voluntary associations, and politics. Actions and transactions in each area are closely examined. The important role of credit institutions and other organizations for sponsoring social change, and the extent to which the founding of such organizations represented an alliance of kinships and business groups are major concerns.

The data used for this study are markedly different from the usual types of social science information. One cannot completely define beforehand the relationships that may exist among individuals or groups. Individuals may be in states (e.g., have jobs, loans, or offices) which can start or stop at any time. More than one state may be occupied at one time. Data sources span various periods. For the study of credit, they include information about social status, economic activity (especially credit transactions with banks such as Cheshire Provident Institution (CPI) of Keene, New Hampshire), and important life events for businessmen and others. Obviously, information is required regarding individuals and groups before and after significant events occurred. Information on all relevant individuals, organizations, and institutions should be continuous and include data on important social relations, such as marriage, business partnerships, and offices held in organizations. Such material are "running" records.

As an example let us examine the complexity of the loan records of the Cheshire Provident. The nature of the physical sources presented several

major obstacles to organizing these records into individual loan records. These problems included the facts that: 1. CPI loan records were entered into a series of ledger volumes, 2. they were not organized by individual, 3. loan indexes by recipients were prepared only twice during the bank's history, 4. one set of five volumes contains most of the information about the loans, 5. another set of 13 volumes contains information about interest payments, 6. assessment of all loans outstanding in 1896 when the CPI was reorganized are contained as documents in a court case in the Cheshire County Court House, 7. during the first 16 years CPI was in business, loan records were included along with records of all other transactions in a completely unsegregated manner in the ledger books. Processing these sources · meant that each loan was treated as a distinct entity. Entries were made whenever the status of the loan changes (i.e., a principal payment was made); thus, any individual might have any number of loans outstanding at any given time. This information, however, was not easily obtained from Cheshire Provident's loan records. In order to accumulate a synopsis of an individual's debt at any given time, it was necessary to go through the loan records and tabulate all those that pertained. Storing the data as running records simply required that this bookkeeping system be handled by the computer. Each payment record represents a record in the data set, only some of which are valid at any given time.

To handle this data we used the following steps:

1. Since the running record sources do not guarantee that all of a set of questions would ever be answered about any case, we adopted · a symbol derivation coding scheme, where each symbol produced elements uniquely identifying each piece of information. For example loan date becomes LOAD, co-signer becomes LCOI. These codes become entries in the symbol table for the FIXFORM program.
2. Using the symbols, the material was coded, either at the archive or using copies in the office on paper. Using printouts of previously entered data as a guide, information was tagged with the identifier and date at this point. It was found that the data coders were quite adept at assigning the correct identifiers in most cases, often relying on intuition or "leaps of historical imagination" which no presently available machine linkage algorithm can duplicate. Most mistaken linkages were caught in later validation steps.
3. The coded material was entered on computer terminals in a free form input file. Given the decreasing cost and increasing power of portable computers in many cases it would be possible to code directly into a computer at an archive thus bypassing this step.
4. The FIXFORM program was run to transform the material into a form which was eventually merged into the master files. The identifiers, dates, and other sort keys were tagged onto each piece of information. FIXFORM is a table driven program and could be and has been adopted for many types of data. All another researcher needs to do is define another set of symbols and enter them in the table.

At the end of this process the data was in a machine readable and uniform structure, yet its original character form also had been preserved. This was essential for later proof reading, editing, recoding, and analysis steps. The output file from FIXFORM were then checked for consistency with the data structure (DUPINVAL), the identifiers were checked against already existing ones (KEYCHK), and the data values were proof-read against the original sources. The file were then either, edited and the process begun over, or merged into the Master File.

Using this procedure more and more files were gradually merged into the Master. It has been found that processing a separate file for each physical data source minimizes mistakes and the time needed to correct them. The "loan file" which includes information on more than 10.000 individual loans, including: type, amount, repayments, cosigners, terms, etc., actually represents a merged file with material from over 35 different physical sources. This file has gradually been augmented by information from a variety of data sources each coded in small files. These include: R.G. Dun reports and published credit ratings for Cheshire County, loan applications, records of loans by other banks in the county, vital statistics and town histories, selected Wills and Probate records, and manufacturing, population, and agricultural census schedules from Cheshire County.

The data which has been compiled is unique in many ways. It was in an accessible data structure maintained by a system of programs that can restructure it at will. The original character literals were preserved. Recoding and retrieval are very easy to do. Use of these methods means that data from "running records" and other sources are linked together by individual so that questions related to community change can be answered.(28)

Industrial Mobilization and Facilities Expansion in World War II

Another example involves industrial mobilization and the expansion of industrial capacity for World War II in the U.S. which is being analyzed by Sweeting. The war greatly altered the relationship between business and government (at least for the duration of the war) with most economic decision making transferred to new wartime agencies in Washington. To staff the new bureaucracies, businessmen were recruited from the private sector to which most returned after the war. Perhaps the most significant change was the government's new role as the banker for industrial expansion. During the war, $28 billion dollars (in wartime dollars) worth of new facilities were built. The federal government directly financed $17 billion of this and accelerated depreciation accounted for $5 billion more. Even the plants built with private capital had to be approved by Washington. After the war, government-built plants were sold on very favorable terms with the wartime operator receiving the right of first refusal of purchase. Many of these plants, and almost all of those financed through accelerated depreciation, had potential utility in the civilian market. The government made reconversion loans available to those who needed them.

The development and implementation of wartime mobilization policies raise many interesting questions for historians and economists, yet they have not received much attention from scholars. Our knowledge of economic change is basically limited to aggregate figures for the amount and number of contracts, the number and value of plants built, the types of goods produced, etc. Our understanding of the war's effect on economic development would be greatly enhanced by analysis of business/government relations on the individual firm level. Because firms interacted with the government in such a wide variety of areas and because government contracting and the type of involvement varied over time for individual firms no simple cross-sectional analysis can capture the full complexity of the mobilization program.

The records for a company level transactional analysis exist. Most of the administrative records of the wartime bureaucracy have been preserved in the National Archives. These include pre-war surveys of available industrial capacity and plans for allocating it during an emergency, contracts, appli-

—

cations for access to raw materials, manpower, and transportation facilities, applications for plant financing, accelerated depreciation and permission to build or expand facilities, applications for exemptions from production restrictions and price fixing regulations, and detailed inventories of machine tools and production technology installed in newly erected and expanded facilities. Also available are materials on contract negotiation, post-war plant disposal and personnel records for some of the new bureaucrats. Some of these records are not available in list form, but instead must be extracted from the minutes of meetings and other narrative sources. All of these sources are example of "running records" which can be exploited by our methodology. Data is being collected from each of the types of records and then organized in a computer data base for each individual firm or even each plant of each firm. Each firm is then linked to data on its pre- and postwar status in the economy. Finally a record is made linking each firm to any of its executives serving in the emergency agencies. The result will be a much more complete understanding of how the war mobilization program affected the economy. This research is currently under way.

The use of our "running records" methodology thus represents a further application of computer techniques to the analysis and organization of historical materials. In an earlier setting we rhetorically asked if the computer can emulate the historian. We do not feel that we have yet reached that point. But we do believe that historians and social scientists can use such powerful methods to begin to address problems that previously were approached in an impressionistic and intuitive manner. Not only do they allow the historian to use narrative data systematically for a wider range of cases, but they also can help to add dynamism and complexity to social scientific analyses. Such a result would mean that scholars would be using the computer as a tool to help solve problems which "traditional" historians have claimed as their own.

NOTES

1 Throughout this paper the term "running records" refers to a specific type of record source. The term "running records methodology" refers to one particular means of utilizing such sources. It is quite possible to use "running records" independently from the methodology described herein. In this paper, unless "methodology is specifically mentioned, we are referring to the type of record.

2 Obviously computers are in use in many areas in history and the social sciences. Mostly, however, these machines are used for complex statistical calculations. As will become apparent, running records methodology calls for the computerization of many more aspects of the research process.

3 In many such approaches, underlying assumptions about the type and the direction of change may exist. Peter H.Smith: Time as a Historical Construct in: Historical Methods XVII (Fall 1984) provides an interesting exegesis of several.

4 See Peter Blau and Otis Dudley Duncan: The American Occupational Structure, New York 1967.

5 See Alice Hanson Jones: The Wealth of a Nation to Be: New York 1980.

6 See Michael Katz: The People of Hamilton, Canada West, Cambridge, Mass. 1973.

7 See David L. Featherman and Robert M. Hauser: Opportunity and Change, New York 1978.

8 Unless one makes some very strong assumptions about the system being in equilibrium at two points in time, comparative cross-sections actually can give little indication about what has caused change, even if they do present the degree to which change has occurred. See William J.Baumol: Economic Dynamics, New York 1970, for example.

9 See Robert W. Fogel and Stanley Engerman: Time on the Cross: Boston 1974, chapter 3.

10 Most econometric techniques have been developed to analyze time series in various situations. See any standard econometrics textbook, such as J.Johnston: Econometrics Methods, New York 1979.

11 The Panel Study of Income Dynamics under the direction of James Morgan at the University of Michigan and the "Parnes" study under the direction of Herbert Parnes at Ohio State University are two of many examples.

12 Nancy Brandon Tuma, Michael T. Hannan and Lyle P. Groeneveld: Dynamic Analysis of Event Histories, in: American Journal of Sociology LXXIV (January 1979), pp. 820-854, and Nancy Brandon Tuma and Michael T. Hannan: Social Dynamics, New York 1985.

13 Sarah Harrison, Charles Jardine, Tim King, Alan Macfarlane, and Jessica Styles: Reconstructing Historical Communities with a Computer, in: Historical Methods XIII (Fall 1980), p. 236.

14 Alan Macfarlane: Reconstructing Historical Communities, Cambridge, England 1978, chapter 2.

15 Macfarlane, p. 37.

16 See Thomas Bender: Community and Social Change in America, New Brunswick, N.J., Rutgers University Press 1978, for a useful survey of the community study literature. Donald H. Parkerson: How Mobile Were Nineteenth-century Americans?, in: Historical Methods XV (Summer 1982), pp. 99-107, casts serious doubt on the view of a highly mobile society in the latter part of the 19th century.

17 R.G. Hutchinson, A.R. Hutchinson and Mabel Newcomer: A Study of Business Mortality: Length of life of Business Enterprises in Poughkeepsie, New York 1843-1936, in: American Economic Review XXVIII (September, 1938), pp. 497-514.

18 Hutchinson (after we recompute the figures to overcome a significant conceptual flaw) shows that, for example, the twenty year persistence rate for businesses is 15.5%. Using our multiple sources the twenty year rate in Keene is 47.0%. Using our single source (R.G. Dun published credit ratings) the twenty year persistence rate is 18.1%. These findings are detailed in Andrew A. Beveridge and George V. Sweeting, Stability, Mobility or Mortality? The Persistence of Business in a Small Industrial Community, 1830-1915, presented at 1983 meeting of the Organization of American Historians, Cincinnati, Ohio, April 1983.

19 Macfarlane, p. 37.
20 Macfarlane, pp. 207-214.
21 T.J. King: The Use of Computers for Storing Records in Historical Research, in: Historical Methods XIV (Spring 1981), pp. 59-64.
22 Much of the work of Philadelphia Social History Project has been collected, in: Historical Methods Newsletter IX (March-June 1976), and Theodore Hershberg ed.: Philadelphia: Work, Space, Family, and Group Experience in the 19th Century. New York: Oxford University Press 1981; the PSHP software is decribed in: Christopher Sobotowski, Software, in: Historical Methods Newsletter IX (March-June 1976), pp. 164-174.
23 Manfred Thaller: Automation on Parnassus: Clio - A Databank Oriented System for Historians, in: Historische Sozialforschung - Historical Social Research XV (1980), pp. 40-65
24 The file design was the creation of Hansen and Ruggles. Many of the original programs were written by Hansen, Ronald Rivest, and David Jefferson. The enhancements were primarily written by George Hess and Nat Eisman under the direction of Beveridge. The SCHEMAT and XREF programs were written by Hess.
25 See Andrew A. Beveridge: Studying Community, Credit, and Change by Using "Running" Records from Historical Sources, in: Historical Methods XIV (Fall 1981), pp. 153-162.
26 For an overview of data base theory see C.J. Date: An Introduction to Data Base Systems. Reading, Mass. 1977.
27 Helena Flam: Beyond Democracy: Work, Credit, and Politics in Paterson, N.J. Between 1890 and 1930: Ph.D. Dissertation, Columbia University, Department of Sociology 1982. Toby Lee Ditz: Ownership and Obligation: Family and Inheritance in Five Connecticut Towns. 1750-1820: Princeton, N.J. 1986 (Forthcoming).
28 For a survey of preliminary findings see Andrew A. Beveridge: Local Lending Practice: Borrowers in a Small Northeastern Industrial City, 1832-1915 in: Journal of Economic History XLV (June 1985).