Developing an Integrated Administrative Database
Draft

Mairéad Reidy, Robert Goerge, Bong Joo Lee

Chapin Hall Center for Children
University of Chicago
1313 East 60th Street
Chicago, Illinois 60637

This is a final draft version of a chapter in the forthcoming volume Exploring Research Methods in Social Policy Research. The volume is edited by M. Little and D. Gordon and to be published by the Ashgate Publishing Company, Aldershot, UK.
©Chapin Hall Center for Children, 1998
Chapter One: Developing an Integrated Administrative Database

Overview

The prevailing wisdom for some time among human service policy makers, providers, and recipients has been that a categorical system of human service provision is an ineffective and inefficient way to support families with multiple and complex needs. When a service delivery system offers discrete specialized categorical programs administered by agencies with varying and often independent mandates and missions, it is not uncommon for families to have multiple caseworkers, each working with different aspects of their needs. In addition, when a system focuses on the individual, family members with similar problems may have different caseworkers. When no single organization or individual takes responsibility for ensuring that services adequately address the full range of family problems in a mutually reinforcing way, service responses can typically fail to facilitate effective, comprehensive, and long-term solutions (for extensive discussion on these issues see, e.g., Agranoff, 1991; Bruner 1991 & 1996; Kinney et al., 1994; Konrad, 1996; Ooms & Owen, 1991(a)(b); the National Commission on Children, 1991).

The debate on how to improve the responsiveness of public social services to children and families, particularly low-income families with multiple, interrelated needs, has focused on the potential strengths of “service integration.” Broadly, service integration is a “process by which two or more entities establish linkages for the purpose of improving outcomes for needy people” (Konrad 1996, p. 5). In short, service integration means linking together services to treat an individual’s or family’s needs in a more coordinated and comprehensive manner (U.S. Department of Health, Education and Welfare [HEW] 1976, p.5). (For a discussion of the potential benefits of service integration see, e.g., General Accounting Office (GAO), 1992; Kagan et al., 1995; Kagan & Neville, 1993; Konrad, 1996; Kusserow, 1991(a).

Interest and confidence in the potential of service integration to improve system responsiveness and deliver more effective and efficient human services have led to three decades of experimentation with service integration initiatives across the United States. Of late, human service reform has gained added momentum, as fiscal constraints and welfare reform hasten the need to improve the self-sufficiency of families with complex problems facing benefit time limits. With states and localities taking the lead, reform that incorporates some degree of service integration is now occurring at some level in most states. The initiatives reflect the growing consensus that in addition to being integrated and comprehensive, services must be family-focused and operated in partnership with communities (for a fuller discussion see, e.g., Agranoff, 1991; Kahn & Kamerman, 1992; Konrad, 1996; Ooms & Owen, 1991(a)(b); Priester, 1996; Waldfogel, 1997). They can range from incremental approaches that enable the current system to operate more effectively to more ambitious strategies that involve agency restructuring, or consolidation, or service delivery model reform (Agranoff, 1991; Kahn & Kamerman, 1992; Kagan & Neville, 1993; Konrad, 1996; Kusserow, 1991(b); Ooms & Owen, 1991(a)(b); Priester, 1996).

Despite widespread interest and confidence in the potential of service integration, the literature suggests that, on average, its success at enhancing system responsiveness for those in need has
been moderate (GAO, 1992; Kagan & Neville, 1993; Kusserow, 1991(a); U. S. Department of Health and Human Services, 1993). Such lack of system responsiveness may, in part, be explained by the absence of adequate information on the service needs of the families they were designed to serve. Service integration initiatives of the past 30 years have not typically been based on a clear understanding of the extent of multiple service among the populations served, although some interesting examples in the literature point to its potential usefulness. Such an understanding can inform service delivery model development, aid in determining the level of resources needed for implementation, and support effective budgeting and resource allocation.

This series of chapters focus on how planning for human service integration can be enhanced by developing and integrating an abundant but often under-used data source; namely agency administrative data records. Administrative data are those records maintained by major U. S. public human service agencies to monitor and track the services they provide. When individual client records are linked longitudinally and across service settings and programs, the integrated database captures the sequence, duration, and movement of children among various agencies and offers a dynamic view of the individual child's experience of service provision. The result is a permanent and continually updatable research tool that has the potential to support a wide range of research. The ability to generate timely and detailed statistical reports on caseload characteristics and trends and to model typical service careers of children are examples of how integrated administrative can inform public administration and public policy. Linked administrative data further supports the analysis of system-wide patterns of service provision and overlap at both the state and smaller geographic areas, and can thus be an central tool in the development, management, and monitoring of new "integrated" service initiatives and programs.

This first chapter continues by focusing on the methods used to integrate administrative data records across agencies and programs. It reviews some of the advantages that administrative data have over sample surveys and other data sources in shedding light on patterns of multiple service use. It outlines the steps involved in accessing and integrating administrative data records across multiple state agencies to produce an unduplicated count of individuals, and families or cases. We focus on acquiring agency data and maintaining data confidentiality, analyzing and documenting source data, designing the database, linking the records of individual clients over time and across agencies, creating summary records representing time intervals, and distinguishing patterns of use at the community and neighborhood level. Chapter Two then illustrates how these data have been used to support service integration reform in the state of Illinois. It presents some of the findings on prevalence and patterns of service use across the human services system at both the state and community level and outlines selected applications of these data in supporting planning for both human service agency consolidation and a new system of integrated service delivery.

**Administrative Data and Its Strengths**

Most major U. S. public human service agencies maintain administrative data records to monitor and track services provided. These records typically contain basic demographic information on each individual served, as well as detailed information about type, dates, duration, location, provider, and cost of services provided. Since the early 1980s, virtually all public agencies have come to rely on computerized information or reporting systems to monitor and track services
provided, to determine the consumption of resources, and to ascertain the capacity to supply services (U.S. Dept. of Health and Human Services, 1991). This administrative data have recently become a valuable source of information for researchers and others who wish to gain a more comprehensive understanding of service use patterns and service pathways of recipients (Goerge, 1997).

The spatial and temporal nature of administrative data lead to many advantages over other data sources for assessing patterns of human service use. Optimally, each agency maintains records of its complete service population across its geographic jurisdiction. The fact that administrative data represent a complete population rather than a sample eliminates the risk of sample bias and improves accuracy of results. Population administrative data further increase the range and complexity of questions that can be investigated, permitting research on low-incidence problems and hard-to-study populations. Available geographical information (such as mailing address, neighborhood or county of residence) can aid in developing profiles of service provision and client composition in a given area or community. Typically, the high costs associated with extensive sampling frames in social surveys render such local-level comparisons extremely costly. Also, because the databases have existed for some time--in many cases since the early 1980s--they possess a historical richness and depth exceeding that of almost any other data. Furthermore, because administrative systems collect data as part of the ongoing administrative process, the costs associated with collecting research data, and the accompanying intrusion, are reduced, if not eliminated. Events are immediately collected instead of waiting for data collection points, as in panel studies, thus reducing the likelihood of errors due to memory loss.

Administrative data have been shown to be more accurate than survey data in certain areas. Although, compared with the average survey, administrative data often have only a small amount of information on each case, the data that are maintained are often of superior quality. Data that people often have trouble remembering in an interview, such as the exact amount of their benefits or the dates on which they received assistance, are more accurately recorded in administrative databases. Brady and Luks (1995), for example, explored the differences between survey responses on the length of welfare spells and administrative data on the receipt of welfare and found a “social desirability” bias in survey responses. Respondents reported shorter spells than recorded in administrative data. Furthermore, because the collection of administrative data is often mandated by law, and because individuals must often be contacted to receive services or aid, certain identifying demographic information (such as name, age, gender, race, and geographic location) is considered to be reliable. In addition, because payments are often made to providers according to length or intensity of service use, the data on dates that individuals entered and exited services may have a built-in check of reliability.

**Building an Integrated Database**

Because the responsibilities of each agency are limited and distinct, each has traditionally developed an information system to track delivery of its own services. A single agency typically does not possess a complete view of an individual or family’s service history or the range of problems with which a person or family struggles. An individual client using more than one service will, therefore, have many records scattered throughout several databases. In order to monitor patterns of service use across agencies and service settings, and over time, it is necessary to integrate, or link, individual administrative records across agencies and services.
Accessing administrative data records and constructing an integrated database are tasks of considerable political, technical, and conceptual complexity. We outline some of the various challenges below.

**Acquiring Agency Data and Maintaining Confidentiality**

The first challenge in developing an integrated administrative database is negotiating access to agency data. Major impediments to access include agencies’ legitimate concerns with confidentiality, increased work burden, and conflict with public responsibilities. Without access to identifying individual-level information, efforts to link individuals across databases will fail. Clients' names are an important (although not the only) element in linking records in an integrated database with a high degree of accuracy (Goerge, Van Voorhis, & Lee, 1994). Agencies must, therefore, be sufficiently comfortable with the established security provisions in order to release client data without expunging identifying information. Furthermore, because agencies have sensitive public responsibilities and typically function with inadequate resources in the midst of close public scrutiny, they can be understandably wary of entering into any arrangement that may create additional work for agency staff or compromise public standing.

It is essential to have in place confidentiality agreements with each agency that meticulously protects clients’ privacy rights and to institute safeguards that maintain confidentiality. Restrictions and stipulations can differ across agencies according to variations in state law and the internal standards agencies maintain with respect to confidentiality (see Solar, Shotton, & Bell, 1993 for a full discussion). In negotiating such agreements, it is important to implement procedures that ensure data security and, once at a research facility, that controls access to data. These procedures include creating an inventory of confidential records when they are received, storing data tapes in a locked facility, maintaining passwords and otherwise protecting files in the database, and training staff in matters of data security. Independent security audits can verify that security measures are sufficient and that staff adheres to them routinely. In addition, although individual-level information is essential to the record-linking process, once complete, most identifying information (e.g., names and addresses) may be moved to a separate file, accessible only to authorized personnel, effectively concealing the identities of clients during database use. Although confidentiality agreements typically stipulate that data can only be reported at the aggregate and never at the individual level, it is also important that further commitment to preventing deductive disclosure of client identities be maintained. Deductive disclosure ensues when an outsider can infer information about an individual from published aggregate statistics based on a sample in which the individual happens to be included (Boruch & Cecil, 1979). Readers should consult Boruch and Cecil and the data-reporting standards of the U. S. Bureau of the Census for procedures to prevent deductive disclosure.

Beyond assuring confidentiality, there must be a clear mutual benefit to sharing data. Agency officials typically recognize that they need better information to improve service delivery, but they often have few resources available for research or data analysis. The prospect of an information resource yielding timely data on client populations and services provided can be attractive. Researchers, however, must be clear about the uses to which data will be put, and those uses must be compatible with the agencies' goals and mandates. Providing agencies with the results of research has became a component of the formal agreements signed by the Chapin
Hall Center for Children with cooperating agencies. Agencies are given the opportunity to review the findings of database research before they are made available to the public. These conditions mutually benefit the research and practice communities, allowing for the expansion of general knowledge while serving the information needs of human service providers.

**Analyzing and Documenting Source Data**

As agency data are acquired, project staff can expect to spend a great deal of time familiarizing themselves with the characteristics of source data. Carefully documenting the possibilities and limitations of each source dataset and assessing comparability of coding across datasets are essential to support and guide later integrated database design (Goerge, Van Voorhis, & Lee, 1994). Documentation involves working collaboratively with agency staff to understand how each data item is defined by the source agency and compiling reference catalogues (written and on-line) documenting every item. Documentation includes variable definitions, value codes, and original data entry rules. Agencies sometimes redefine variables or discontinue their use because of changes in operations or agency mandates, and documentation should reflect any such alteration and its effective dates. Other pertinent information about the service context in which data were collected, including, for example, agency protocols for tracking clients and opening and closing of cases, should be documented. This information is not generally included in the data guides and reports of human service agencies. Although each administrative dataset is documented to a varying extent, much of it must be obtained by interviewing agency staff.

Information about how “good” the data are is seldom documented, and determining the quality of the data is not a research activity that has any clear protocol. The task can be either active or passive. An active approach involves checking the data with the object to which the data belong. This could be an individual, a service recipient or agency staff member, or a paper record that is the definitive source of the data record or item, such as agency data entry forms. A passive data quality review involves comparisons with other sources of similar data. Active data quality review is seldom done because of cost and feasibility; it is often impossible to find a service recipient after the data are collected. However, there are instances of agencies that do have data quality assurance units that conduct this type of work. Knowing when such units exist can enhance confidence in the quality of administrative data.

Again, interviewing those who maintain the database, use the data, or train those who will collect the data is generally the most effective source of information about data quality. However, finding in a bureaucracy the individuals who actually know about all the fields in a database can be an iterative process. The types of data most germane to each agency’s functioning tend to be the most used and, therefore, the most reliable. Because individuals must usually be contacted in the normal course of providing service or aid, identifying information, such as names, addresses, and social security numbers, are usually accurate and maintained over the period during which the individual is in contact with the organization. This information allows one to update service records, track individuals and families over time, and is essential to linking records to other administrative databases (described in detail below). Finally, it should be noted that learning about the reliability, validity, and accuracy of data may only be possible after the data have been analyzed. For example, there may be dramatic variation among counties in a
state, which may not be understood until the data are actually analyzed and experts can respond to the results.

Carefully assessing which data are truly comparable across databases is essential in integrating data. Agencies are not bound to any one standard in designing their administrative databases. Definitions of even such basic fields as race or ethnicity may vary from one agency to the next, to say nothing of more complicated variables, such as handicapping conditions. The researchers must, therefore, standardize variables. In addition, duplicated or redundant information should be eliminated, unreliable data archived, variables not currently needed should be suppressed, and needlessly elaborate variables (with large numbers of values or significant length) should be reduced.

**Designing the Database: Relational Models of Integrated Administrative Data**

The process of turning raw administrative data into an integrated database that can be used for policy and program-relevant research must be based on a conceptual model of how various data elements are related to one another. The data received from administrative systems, typically derived from nonrelational databases, contain much redundant information and are often poorly structured for anything beyond the specific reporting requirements of the database (Goerge, Van Voorhis, & Lee, 1994). For example, within a state, each agency might have its own separate client database containing information about the clients, services provided, and family members. Simple aggregation of these separate databases without a clear conceptual data model will result in massive redundancy of information (since a client can have many “client records” across these agencies), incomplete information (various services a client received from different agencies will not be appropriately linked to the client), and sometimes confusing data (a person recorded as a client in one agency may be recorded as the father of a client in another agency). The family or case unit may be defined in different ways across agencies, and even within agency systems it may be defined differently across programs. For example, the case unit in one agency’s database may be defined to include everyone living in the same household, and in another database, case unit may refer to the grantee and dependent children. Principal objectives in designing a relational database model must be to preserve the universe of relationships, actors, and events implied in the original data, while minimizing storage costs and maximizing speed of retrieval. One must also produce a general structure of the interrelationships among the data elements that can order a massive quantity of raw data into a unified database.

These objectives are achieved by designing a database using a mixture of object-oriented and relational techniques. Using a relational data modeling approach, one can create a single data file that contains an *unduplicated* list of people identified across agencies and a single data file that contains a list of agencies and each agency’s services. These two files can then be *related* to each other, specifying the appropriate service relationships: The result is a set of clients of an agency and a set of agencies providing services to a client. As in the example provided above, another important relationship to represent is the family membership, which essentially involves linking individuals within the people data file and specifying how individuals are related to one another; for example, father-child, mother-child, siblings, and so forth. In order to define these family relationships as effectively as possible, it is important to use available and complementary case-level information across programs and agencies over time. Hence, when
there are multiple individuals (adults and children, whether active or inactive, at any point in
time) associated with one case unit, those individuals are collated to one case. When additional
individuals related to these cases are found across other programs, they are combined into these
cases. The resulting relational database is more storage efficient, and unified (because redundant
information has been eliminated) and offers a clear and complete view of the system.

Concentrating on objects (e.g., actors and events) rather than on specific records or variables
yields a general structure capable of ordering and subordinating the massive quantity of data and
the wide range of variables from the constituent databases into an integrated database. This
general, object-oriented design is illustrated using modified entity-relationship diagrams that are
implemented in a relational database system (See Brookshire, 1993; Date, 1990). An Entity-
Relationship Diagram (ERD) is a visual representation of the structure of a database model.
There are three types objects in an ERD: entities, attributes of entities, and relationships
between entities. An entity is an object of interest in the subject area being modeled. Entities can
be people or organizations, events or episodes, abstract concepts or collections. Attributes are
the pieces of information that describe entities, such as a name, a date, or a location. An
instance of an entity has specific values for its attributes. For example, one instance of a child
might be John Smith, born 1/1/96 in Cleveland, Ohio.

Consider the example of a model developed to monitor a simple foster care tracking system.
(See Figure 1). In this example, we are interested in tracking children within foster care cases as
they move from placement to placement and in maintaining some information on the head of
household in the child’s family. Boxes represent the entity types, and the lines between them
represent the relationships. The notation at the ends of the relationship lines provides
information about two properties of the relationship: cardinality and optionality. The cardinality
of a relationship, indicating either one instance of the entity or many, is represented by a bar or a
crowsfoot, respectively. The optionality, represented by a circle or a bar on the line, indicates
whether the relationship can be optional or not. Attributes have been omitted from this diagram,
but normally would be shown within the entity boxes. When interpreting the diagram, the
relationship names should be read from top to bottom or from left to right. Thus, Zero or one
Head of Household is related to one or more child[ren], or each provider takes care of one or
more living arrangement, each child is part of one or more cases, and each case can include one
or more child, and each child has resided at one or more living arrangement.

The great advantage of this structure is that it allows researchers enormous latitude in how they
structure their analyses and in what they determine to be the unit of analysis. A particular set of
children, families, events, geographic areas, or dates may serve to define the base population for
a given analysis. This initial transformation of the data into a relational structure must be carried
out for each source database.

Once the multiple source databases have been transformed into relational structures and are
linked together (see below for linking the records of individuals), the integrated database design
features a primary set of tables (or entity types) that describe the main actors and events in the
data. Additional tables detail the varying information for different types of actors and events
from each of the databases. Figure 2 shows that the primary individual table, for example, is
associated with separate tables for children, parents, and foster parents. Likewise, the primary
event table links to a number of tables for different events, such as case openings, foster care placements, or case closings (mental health, foster care, or public aid). In addition, tables may be created to group together data on geographic or governmental entities (e.g., communities or counties), or individuals (e.g., families), or dates (e.g., months or years).

Figure 1 Entity relationship diagram illustrating the foster care tracking system.

Figure 2 Schematic of object relationships in an integrated database across three data systems (Mental Health, Foster Care, or Public Aid)

**Linking the Records of Individual Clients Over Time and Across Agencies**

Once the relational structures are established, the primary technical challenge in creating an integrated database from diverse databases is accurately linking the records for a specific person across the different agencies’ databases. Normally, there is no single field or variable found in
all databases that can reliably establish the identity of a person across different databases and agencies.

In a single database, two records are considered “linked” when they both contain the same value for a specific field, generally called the person identification (ID) or client ID. This ID should uniquely identify the person. No two persons should have the same ID, and every record associated with a given person should have the same ID. Every agency uses its own system for assigning a person an ID. Sometimes the ID may even be different across one agency’s programs. In a number of cases, a single agency may issue a person more than one ID. A new ID may be assigned each time a case is opened or each time a person receives a service. It is also possible for a new ID to be assigned because of an agency’s inability to associate a person with his or her previous ID.

A person’s social security number (SSN) is frequently considered such an ID because a field for SSN is often included in a person’s main record. However, there are problems with using a SSN as the link ID. It may be missing or otherwise unavailable in one or another database. And data entry, transcription, or other errors can create false links or scramble a true link. Other fields that might be used to establish a link are equally problematic. For example, names or birth dates that match exactly can easily refer to two different persons (Goerge, Van Voorhis, & Lee, 1994).

The most reliable means of matching records has proved to be a process called probabilistic record linkage, first developed by researchers in the fields of demography and epidemiology (Newcombe, 1988; Winkler, 1988; Jaro, 1985, 1989; Baldwin, Acheson, & Graham, 1987). Probabilistic record linkage assumes that no exact match between fields common to the source databases will link a person with complete confidence. Instead, probabilistic record linkage calculates the likelihood that two records belong to the same person by matching together as many pieces of identifying information as possible from each database. First and last name, birth date, gender, race, SSN, address, and county of residence are commonly used matching fields. Other fields are used if they are found in both databases and are useful in distinguishing one person from another, for example, mother’s name, previous name or address, the date of a comparable event. When multiple pieces of identifying information from two databases are comparable, the probability of a correct match is increased. Probabilistic record matching, therefore, weighs a number of identifying items in the two records of a proposed match in order to arrive at a probability that they refer to the same individual. Matches with high probabilities are deemed identical.

The minimum number of fields needed for linking varies with the size of the population and the confidence level desired. In general, more fields provide a higher level of confidence. In a small population, one might achieve good links by matching only name and birth date. In a larger population, however, additional fields must be matched to create links with the minimum confidence desired. Once a match has been determined, a unique number is assigned to the matched records so that each record can be uniquely identified. The end result of computer matching is what is referred to as a master “link-file,” which contains the unique number assigned during matching, the individual’s identifying data (name, birthdate, race or ethnicity, gender, and country of residence), and all the identification numbers assigned by agencies from which the person received service. Thus, when any individual has had contact with more than
one agency, the various records belonging to the individual are linked, resulting in a particularly rich array of information in these cases.

Multiple sources of data and record-linkage techniques make it possible to detect error and to correct it in some cases. The demographic information of an individual who has received two or more services, for example, can be checked using multiple data sources. Some agencies collect the same demographic information on each client year after year, and thus consecutive records may be compared to detect one-time errors. Where multiple records exist for a client, some errors, such as incorrect racial coding, may be cautiously corrected.

Creating Summary Records Representing Time Intervals

Returning to our example of foster care tracking, once all the records pertaining to an individual child have been linked, new records may be created to represent significant periods in the child's service history and to express other facts that may be inferred from given data. The new variables, or summary records, are produced by manipulating or combining existing variables according to mathematical formulae. The most important summary records for our purposes are records expressing time or duration, such as the duration of a child's eligibility for Medicaid or duration of foster care placements.

Duration is expressed as a "spell," an interval bracketed by two events occurring at different times and consisting of the elapsed time between them (Bane & Ellwood, 1986). The event dates and personal identifiers contained in administrative data provide the necessary elements for converting event data into spell data. Roughly, this conversion is accomplished by identifying all event records associated with a particular child, sorting them into chronological order, and creating a new data record for each between-event interval. Records constructed in this way are entirely new analytic entities in which the time interval replaces the event as the unit of analysis.

Summary records allow for the kind of simplified statistical analyses familiar to social scientists. For example, summary records may be used to analyze patterns of service provision and use within a single agency, or to assess the relationships of use among several agencies. We at Chapin Hall have used summary records to study patterns of duration and re-entry in foster care (Goerge, 1990; Wulczyn & Goerge, 1992). The records may also be used to analyze the timing of service interventions and long-term patterns of service use.

Aggregating Data to Appropriate Geographic Level of Analysis

Among the strongest impulses in human services reform today is the desire to root services more firmly in local communities. Critics of the "fragmented" categorical service system dream of replacing it with more localized and integrated community-based systems. Others insist merely that effectively designed services, whether categorical or not, must somehow take into account environmental factors and conjunctions, that is, must be sensitive to the "social matrix."

When addresses exist in the database, the multiservice database can supply a richly detailed foundation for exploring these ideas. Most of the records in our Illinois database contain some kind of geographical information (such as a mailing address, neighborhood, or county of
This information is then used in a process known as “geocoding” involving the conversion of address information into coordinates of latitude and longitude, facilitating the mapping of points, lines, or regions on a map. When records are “geocoded,” the result is a profile of service provision and client composition in a given area, geographic level, or community. For data to be a useful tool for planning, implementing, and evaluating policies and programs at the community level, it should be presented in a format that corresponds to boundaries of a particular area (such boundaries could be the entire area of a community or its subareas). When demarcations (whether actual street addresses, census tracts, or census blocks) include detailed geographic information, we can combine all the information for a certain area of interest in order to meet a community’s particular information needs. Once the geographic “boundaries” of the area are defined, the geocoded information can be specifically tailored for that area. Individual-level data can thus be aggregated into census blocks and tracts, community or neighborhood areas.

Once addresses are converted into the geographic information, key community information can be integrated with the geographic information into a single database to provide much richer, and more accessible, view of communities. For example, once clients’ addresses are geocoded and located on a map, detailed socioeconomic and demographic information of the community such as education levels, income levels, or employment rates can be attached to the service information. This final set of information is an integrated database that contains the spatial relationship between the actual patterns of services and the profile of the communities served by the agencies.

**Conclusion**

This chapter has introduced the strengths of administrative data and outlined the steps involved in building an integrated administrative database and identifying an unduplicated count of individuals, and families or cases across multiple state agencies and programs. We have discussed the challenges involved in negotiating access to and maintaining confidential agency data, and set out guidelines on how to analyze and document source data. Our discussion of designing the database has introduced the concepts of relational data structures and entity-relationship diagrams, and outlined the strength of such structures in allowing researchers significant latitude in how analyses are structured and in the choice of unit of analysis. Linking the records of individual clients over time and across agencies has been presented as the primary technical challenge in creating an integrated database, and we have outlined the process of probabilistic record linkage that we consider to be the most reliable means of matching. The creation of summary records representing time intervals has been outlined, and their ability to facilitate a wide range of simplified statistical analyses of duration and sequence of service use discussed. Finally, we have focused on the process of geocoding to distinguish patterns of use at the community and neighborhood level.

The ability of administrative data and integrated databases to accurately document system-wide patterns of service use across programs, over time, and at the state and smaller geographic areas makes them essential tools in planning and evaluating human service integration initiatives. We turn in the next chapter to how integrated administrative data has been used to support planning
for both human service agency consolidation and a new system of integrated service delivery in the state of Illinois.
References


Endnotes

---

1. Konrad states that this broad and inclusive definition is similar to those proposed by the National Center for Service Integration, the National Center for Children in Poverty, and the U.S. General Accounting Office.

2. Examination of service integration initiatives across the country highlight some interesting examples of when multiple use analyses were successfully used to plan and implement reform. In order to determine the extent to which greater coordination with schools was possible and necessary, the highly publicized and extensive feasibility study carried out by the New Beginnings Project in San Diego included an extensive assessment of the extent of multiple service use among families in a low-income, multi-ethnic elementary school (New Beginnings 1990, Cohen and Ooms 1993). A similar assessment of the prevalence of service use among students attending several Oakland schools across 19 programs was conducted to assess community needs and resources (Cohen and Ooms).