RESHAPING ADMINISTRATIVE DATABASES INTO TOOLS FOR POLICY-RELEVANT RESEARCH

Early Childhood DataCONNections

An Inventory and Analysis of State Databases on Early Childhood

July 2003

The Child Health and Development
Institute of Connecticut, Inc.

THIS REPORT WAS AUTHORED BY:

Frances Duran & Susan Wilson
Child Health and Development Institute

David Carroll
Words & Numbers Research, Inc.

BASED ON RESEARCH AND ANALYSIS BY:

David Carroll
Words & Numbers Research, Inc.

About the Early Childhood DataCONNections Project

Early Childhood DataCONNections is a public-private partnership of the State of Connecticut's Department of Social Services (DSS) and the Child Health and Development Institute of Connecticut (CHDI). The project mission is to promote well-informed decisions on policies and programs for young children by improving state agencies' research capability. As part of this effort, DataCONNections is bringing together state agency staff, researchers, community advocates, service providers and legislators to identify and address some of the needs for better information on key early childhood issues.

About the Child Health and Development Institute

The Child Health and Development Institute of Connecticut is a not-for-profit organization established to promote and maximize the healthy physical, behavioral, emotional, cognitive and social development of children throughout Connecticut. CHDI creates, supports and facilitates innovative primary and preventative strategies for children, and works to maximize the effectiveness of the institutions and systems that contribute to their well-being.

Acknowledgements

DataCONNections is grateful to those who dedicated time and expertise to the completion of this report, especially members of the DataCONNections State Steering Committee and staff interviewed for the study, listed in Appendix C.

This report was made possible through generous support of the Early Childhood DataCONNections project provided by:

Child Care Bureau of the U.S. Department of Health and Human Services
Children's Fund of Connecticut
Connecticut Department of Social Services
David and Lucile Packard Foundation

For additional copies of this report, call 860-679-1519 or visit www.chdi.org

Executive Summary

Overview

In Connecticut, as in most states, research in early childhood is limited. As a result, policymakers, administrators and others are frequently in the position of making critical decisions affecting young children and their families without adequate information. Early Childhood DataCONNections, a project of the Child Health and Development Institute of Connecticut (CHDI), was created to help address this research gap. The project's mission is to promote sound early childhood policy by improving the quality of information on Connecticut's young children and their families through research, data analysis, data development and information sharing.

As part of this effort, DataCONNections conducted a scan and analysis of administrative databases used by Connecticut state agencies for managing programs and aggregating data on child populations and services for children and families. As a result of this study, DataCONNections has created a compendium of state agency databases that record early childhood data. In addition, DataCONNections has developed a set of recommendations for enhancing these databases to accommodate policy-relevant research that will help guide the development and improvement of policies affecting young children and their families. These recommendations, coupled with general findings on administrative databases stemming from the analysis, are presented in this report.

Based on the criteria outlined above, over 80 databases from the following state departments and auxiliary agencies were identified and reviewed:

- Children and Families
- Education
- Labor
- Mental Retardation
- Public Health
- Social Services
- Community Colleges (Connecticut Charts-a-Course)
- United Way of Connecticut (Infoline)

Profiles of these databases are available through the Early Childhood Data Source Clearinghouse. See below for more information.

Early Childhood Data Source Clearinghouse

DataCONNections has established a relatively comprehensive compilation of data sources on young children in Connecticut through the inventory of state administrative databases discussed in this report and a scan of published data and internet resources. These data resources are the basis of the Early Childhood Data Source Clearinghouse.

The Clearinghouse includes profiles and supporting documents on the state administrative databases, which are available through Sue Wilson at info@chdi.org. It also includes a searchable database of published and internet resources on early childhood issues such as health, early learning, safety and economic security. The database is available on-line at www.chdi.org, under the Resources section, in a user-friendly format that allows visitors to query the database directly.

Purpose

This document is primarily a tool for state agencies and others interested in expanding the utility of state administrative data to support policy-relevant research. The findings and recommendations outlined in the report are intended to help users assess the limitations of their own databases and develop strategies for overcoming them. Those in the process of developing new systems are also encouraged to refer to this document – especially the promising practices sections - during the planning stages.

However, this document is not meant to be simply an assessment tool or a compendium of state agency innovations. It is meant to spotlight the untapped research potential of state administrative databases and to inspire changes that will begin to unleash this potential. These databases house a wealth of information that, if properly harnessed, analyzed and shared, could provide valuable insights for policymakers and others making decisions affecting young children. Through this report, and a host of other activities, DataCONNections hopes to guide this vision for enhanced state research capacity and make it a shared vision for all those interested in promoting the well-being of children in Connecticut including, but not limited to:

- State agencies
- Researchers
- Policymakers
- Family court officials
- Advocates
- Families

Key Findings

- Most state agencies have not focused on their needs for new research or analysis of existing internal data nor on the resources and infrastructure required for such efforts.
- Since most state agency databases have been established primarily for administration, case management, reporting and accountability, they are difficult to use for research purposes.
- State agency databases do not use common record identifiers. As a result, one client could have multiple identification numbers across agencies or even within an agency.
- Variables are coded differently across agency databases and in restrictive formats, hindering the ability to link databases and conduct comparative analyses.
- Some of the data collected is very sensitive, requiring strict measures to ensure privacy and confidentiality.
- In some cases, data on key variables is not collected, inhibiting analysis of critical policy questions.
- Some data fields contain invalid entries that can skew results.
- Several programs lack the staff or equipment to automate data collection.
- Some systems automatically overwrite data with no provisions for archiving it for future reference.

Recommendations

- Build a foundation of commitment from state agencies and other key stakeholders (e.g. legislators, who mandate data collection) for developing databases that support policy-relevant research.
- Develop sample formats for formal interagency agreements among state agencies and between state agencies and outside researchers to advance the goal of producing policy-relevant research.
- Develop protocols and institute an Institutional Review Board (IRB) or its equivalent in each agency to facilitate external access to agency database information and to stimulate research by outside parties using state databases.
- Train and educate agency staff so that they have the skills and vision to expand the functionality of current databases.
- Automate data collection where paper filing is currently in use and ensure that all data collection systems archive data.
- Develop guidelines and standards for constructing and modifying administrative databases to facilitate research and linking data (e.g. common record identifiers and standardized coding of data).
- Initiate several pilot projects focused on enhancing and linking databases and circulate information to other agencies on best practices.
- Develop partnerships between state agencies and academic researchers and showcase positive outcomes from such collaborations.
- Ensure that all audiences benefit from state data by publishing Internet-accessible reports in a timely manner and in a reader-friendly format.

Conclusion

Recognizing that sound policy requires reliable information, Early Childhood DataCONNections is partnering with state departments to help them realize their research and data analysis potential. Mindful of the focused purposes for which state administrative databases are created and used, as well as the significant opportunities they offer for policy research, DataCONNections is working diligently to increase the utility of these untapped resources. This report is a pivotal piece of this effort, identifying needed adjustments to enable administrative databases to support policy research. However, for state agencies to fully maximize their research potential, agency culture must shift toward valuing and raising expectations for policy-relevant research. Furthermore, linkages between the state and the external research community must be pursued.

Toward these ends, DataCONNections will continue to work with state agencies, using this report as a framework to provide an individualized assessment of the strengths and weaknesses of each agency's databases and to discuss their potential for research. In addition, through research roundtables and other information sharing methods, DataCONNections is engaging the research community and fostering state agency/research partnerships.

Introduction

As with many other areas of public policy, early childhood research at the state level is limited. Questions such as how are young children faring in Connecticut, how effective are services to young children and their families, and how do public policies affect the well-being of these children remain difficult to answer. Recognizing this research gap, Early Childhood DataCONNections, a public-private partnership of the Connecticut Department of Social Services and the Child Health and Development Institute of Connecticut, was formed.

A critical goal of the Early Childhood DataCONNections initiative is to promote better information on the status of Connecticut's young children, birth to age eight, so that effective early childhood policies can be implemented. Such policies should be tailored to the unique needs of young children; designed to address any unmet needs; and shaped by reliable information on what does and does not work.

In order to accomplish this goal, a multi-faceted approach is necessary. In partnership with state agency staff, researchers, community advocates and legislators, DataCONNections has developed and is implementing a scope of work that includes:

- Building a strong infrastructure of sustainable research tools and processes
- Creating and carrying out an early childhood research agenda
- Providing research services and consultation to the state
- Linking the research community with state government
- Raising awareness of the value of and potential for policy-relevant data collection and research
- Developing and promoting the use of new early childhood indicators

To begin building a strong foundation for early childhood research, DataCONNections conducted a scan of over 80 administrative databases used by Connecticut state agencies for managing programs and aggregating data on child populations and services for children and families. Although child data is also available through various federal sources (e.g. Centers for Disease Control, Head Start Bureau), this scan focused solely on state agency data. The participating agencies include: the State Departments of Children and Families, Education, Labor, Mental Retardation, Public Health, and Social Services as well as Community Colleges (Connecticut Charts-a-Course) and the United Way of Connecticut (Infoline).

The resources developed from this scan include a descriptive compilation of databases that record early childhood information, listings of variables included in each database, as well as a library of codebooks, data collection forms, computer screen printouts, data dictionaries and generated reports. These tools will assist state agencies, researchers and others in examining what data are being collected and by whom, and will help them to identify opportunities for new research.

Another valuable product gleaned from this endeavor is a summary of limitations inherent in current administrative databases that create barriers for conducting policy research. These findings are outlined in the first part of this report. Part II addresses practical steps to overcome these limitations and expand the functionality of administrative databases, as well as ways to improve state agencies' overall collection, analysis and dissemination of data. Throughout the report, promising practices from a variety of state agencies are highlighted.

These examples illustrate efforts of varying scale that are taking place within and across agencies that directly or indirectly support policy research.

Key Terms

The following terms will be referenced throughout this report. To ensure a common understanding, these terms are defined below.

Policy or Policy-relevant research: Research that informs the development and improvement of public functions such as regulation and enforcement, education and public awareness, direct services and financing of services.

- It includes **baseline and trend information** that describes the status, condition or characteristics of a population, in this case, young children and their families or the providers of services.
- It also includes **program performance information** that may demonstrate the coverage of the population by the service, the referrals, intake, type of services, cost, providers and proportion of eligible clients who use the service.   Most state reporting on services falls into this category.
- A third layer of policy-relevant research produces **evaluative information**. It is concerned with how well services meet their objectives. Outcomes attributed to the services are examined. Coordination and transitions between services and equity and quality are important to evaluative research. At its most complex, this research examines outcomes and attributes them to multiple conditions and interventions.

Administrative data: Administrative data is collected by state agencies and used for:
- record-keeping and case management;
- monitoring and evaluating program performance; and
- ensuring agency accountability.

PART I

State Administrative Databases: Opportunities and Limitations

Before meaningful policy research using administrative data can take place, it is essential to identify and address existing barriers. One of the most immediate obstacles is competing agency demands and responsibilities. State agencies' first priority is client service delivery. Relative to this complex task, research is given low prioritization. In fact, most state agencies have invested very little time and resources toward research planning and development.

Other major barriers are the current limitations inherent in state administrative databases. The following section highlights some of these shortcomings.

Findings

1. **Currently, administrative databases are primarily designed for case management and accountability, making them difficult to use for research purposes.**
   Many agency databases have a great deal of both qualitative and quantitative data. This can be attributed to the fact that they are primarily used for administration, case management, reporting and, to some extent, accountability. In many instances, data is captured in narrative format rather than in defined fields or categories, which presents a barrier to systematic analysis.

2. **Databases do not use common record identifiers.**
   Identification numbers (alpha or numerical), which are needed to identify cases for aggregating and tracking data, vary across databases. Some use Social Security numbers; other use computer generated numbers; and others use their own composite numbers. As a result, clients can have multiple identification numbers across agencies or even within the same agency. This impedes the linking of databases across agencies and the matching of client records across programs.

3. **Variables are coded differently across databases.**
   The ability to compare demographic variables across agency lines is another factor that can either enhance or diminish the research potential of administrative databases. Databases in agency settings are designed to manage and monitor program operations and variables are defined in a manner that relates to the individual program's needs or conforms to the various mandated state and federal reporting requirements. Some examples of coding variability are highlighted below.

   a. **Variability in the number of categories.**
      The number of categories for the variable "ethnicity" ranges from over twenty in one database to the five categories that many agencies use: Asian/Pacific Islander, Black/African, Hispanic/Latin, American Indian, White/Caucasian.

   b. **Alpha versus numeric codes.**
      Some databases code the variable "gender" or "sex" as M and F, while others use the numeric values 1 and 2.

c. **Diversity in coding schemes.**
Databases use several different formats to capture information as evident in the variable "birth date." Not only does the sequencing of day, month and year vary, but also there is variability in the number of digits assigned. For example, January 3, 1953 might be coded as:

- 1 3 1953
- 01 03 1953
- 03 01 1953
- 1953 01 03

4. **Coding of variables is not conducive to research.**
Some variables are collected in formats that limit statistical computation and analysis, essentially leading to the loss of valuable information. One restrictive data collection method is the use of numeric range categories as opposed to actual numeric values. This practice is common for variables such as age, years of education, service, income, cost and others. As the following chart shows, a person 16 years old would be entered into the database as a "2" instead of the actual age and could not be differentiated from a 17 year old.

Category	Ranges
1	14-15
2	16-17
3	18
4	19-21
5	22 and above

Agencies will also record affirmative or negative responses without including details that would be useful for research and planning. For example, a data field on whether a child has participated in a preschool education program might be coded for "yes" or "no" when the type of program would be a much more useful designation. In another instance, the field might specify the type of program, but not be able to accommodate information on children who participated in more than one program.

5. **Databases capture sensitive information that requires strict attention to privacy and confidentiality.**
When databases are used for case management, sensitive data is collected, raising valid concerns about privacy and confidentiality when exploring use of these databases for research purposes. Three of the agencies surveyed have Institutional Review Boards (IRBs) in place to determine whether and how client level data can be used. The other agencies have no formal body or guidelines for review.

6. **Data on key variables is not always collected.**
In some cases, basic but critical data is not collected, leaving gaps in information. As a result, important questions with policy implications remain unanswered.

7. **Data in some fields is recorded erroneously.**
Because administrative databases are primarily tools for service workers and managers, the accuracy and amount of data they contain is directly related to the usefulness of that data in day-to-day operations. In cases where the data is irrelevant to the immediate

task, it is common for busy workers to skip over a data field entirely or, if data entry is required to proceed to the next field, they will often enter a "dummy" value (e.g. $1.00 for a full-time worker's annual salary.) An unsuspecting researcher, internal or external to the department, might not know that some data could be skewed because of these erroneous values. Researchers should explore the possible unreliability of certain fields when evaluating and cleaning a dataset.

8. **Not all databases are automated.**
   Some programs collect data, but resort to filing paper forms in filing cabinets because they are without even a simple electronic database or lack resources for data entry. Although these programs amass important information, without an automated system, access to and utility of this data is severely limited.

9. **Not all databases archive information.**
   When updating database fields with new information, some systems simply overwrite the old information without archiving it. This compromises the ability to conduct comparative analyses and longitudinal studies.

PART II

Converting Administrative Databases into Tools for Policy-Relevant Research

Many questions arise for those trying to shape effective public policy for young children and their families. Who is served and who is under-served? How can services be targeted to those in need? How effective are specific interventions? What are successful programmatic outcomes? What services are cost effective? Are families able to navigate the system and manage transitions from one program to another? Administrative databases can answer these questions in part or in whole, but their development must be considered with a new function in mind: *How can the database be utilized to accomplish multiple goals - accountability, case management and policy-relevant research?* This section of the report outlines recommendations and, where appropriate, practical steps for making this transition and optimizing the functionality of administrative data.

Recommendations

1. **Secure commitment from state agencies and other key stakeholders for enhancing databases to accommodate policy research needs.**
   Without support, particularly from agency leadership, making changes or enhancements is unlikely. To build commitment and motivation, agency staff, policymakers and other stakeholders need to be made aware of the value of policy-relevant research, where opportunities for such research exist, and how to create an information rich environment for policy development.

Promising Practice: An internal research committee sanctioned by agency leadership

The Department of Children and Families established the Connecticut Consortium for Applied Child Welfare Research to provide advice and consultation regarding a department Research Agenda and to find ways to encourage partnerships with researchers from colleges and universities. The Consortium was created at the request of the Commissioner. It is active in promoting these linkages in very practical ways, e.g. model agreements, the Research Agenda, small grants to researchers, a web site, conferences and internships.

2. **Develop sample formal interagency agreements to assist in linking and interfacing databases.**
   Formal data sharing agreements among state agencies allow the parties to be clear and comfortable about their respective expectations in the enterprise. Such agreements result from negotiation of each agency's interests. To facilitate this process, it is important to identify what eases and inhibits current efforts to share administrative data between agencies. In addition, one must address the administrative or regulatory issues that affect interagency data sharing and identify the key elements of existing formal interagency agreements. A review of data sharing agreements that have been used in Connecticut and other states should be undertaken to inform the drafting of recommended agreements.

3. Develop mechanisms to facilitate external access to agency database information and to stimulate research using state databases.

The efforts of outside research organizations to utilize state agency databases for research purposes should be both encouraged and facilitated. Research initiated by academic institutions, policy development organizations and advocacy groups have much to offer the state government with regard to assessing and improving service delivery. Furthermore, linking state agency research needs with researchers' interests and capabilities can be an efficient way to meet these needs, particularly when internal research capacity is limited.

However, it is critical that these groups are aware of state databases that are available and accessible, and under what specific conditions. Having guidelines in place that ensure the privacy and security of clients *and* provide for researcher accountability should enable state agencies to demonstrate openness that encourages other state agencies and outside organizations to conduct research using state databases.

Below are two recommended strategies for achieving these goals.

a. Create a research protocol.

A recommended protocol between state agencies and research organizations would engender confidence and trust, as well as clarity among agencies and outside researchers about what can and cannot be done. It should address: sharing reporting forms; physically accessing data; electronically accessing data; protecting client privacy and confidentiality; usage of data once accessed; ownership of analysis; and, agency review and reporting of results. Because state agencies differ in their relationships to their clients and have different state and federal constraints, a single protocol that all agencies would use is unlikely. The protocol could therefore present optional language for different agencies on some topics.

b. Establish Institutional Review Boards.

Central to increased utilization of state databases is an Institutional Review Board (IRB) or its equivalent that will implement the agency policy. Each state agency should have an IRB with clearly written guidelines that address:

- the control the agency will maintain over use of their data;
- the steps that must be followed to obtain permission to use the data;
- what exactly can be done utilizing the data;
- what information researchers can disclose about the state agency operations; and,
- the protocols that will be used to protect the rights, privacy and welfare of human research subjects.

Also, agencies need to examine under what conditions researchers will be granted access to individual client data. Aggregated data can only be described; it cannot be used for in-depth statistical analysis. The role and authority of the state agency in making available individual data that is collected and stored at the local level should be addressed.

c. Disseminate information to the research community on state databases and state research needs.

In order to facilitate policy research partnerships, state agencies should share descriptions of their databases, including conditions of the data, and encourage researchers to use them. In addition, state agencies should articulate their research needs by developing a research agenda that they can circulate within the research community.

Promising Practices: Facilitating Possible Policy Research Partnerships through Data Sharing

Conducting a thorough inventory of agency databases

The Department of Public Health (DPH), Bureau of Community Health publishes a compendium of databases that provides a profile of each database. For each database, it summarizes data collection methodology, the unit of analysis (e.g. individual, family), the population covered, the availability of key variables (e.g. race, region), recent reports and the agency contact. This metadata is helpful to department and community users to determine what data may be available, the format and how to access it. This is an important first step to building linkages and encouraging analysis.

Developing a data warehouse

The Department of Social Services (DSS) is preparing the first large state agency data warehouse. It will archive extracts of the Eligibility Management System (EMS), which is the state administrative system for public assistance, Medicaid, Food Stamps and some smaller programs. Over time, other databases will be added in usable formats. Without this warehousing mechanism, DSS has for over 13 years had to program ad hoc reports or create new report formats whenever it needed to query the massive mainframe system. Simple desktop programs will be used to produce these reports once the data warehouse is in place, making it easier to accommodate requests for data.

Consolidating databases and streamlining data

The Department of Children and Families' (DCF) Bureau of Health, Medicine and Education is consolidating 25 behavioral health services databases into a comprehensive data collection system and warehouse. The Behavioral Health Services Database collects several types of variables including demographic, service dates, service/process, assessment (pre and post), outcomes and billing. Contracted service vendors are required to provide these data monthly to DCF, coded in a uniform format and transmitted electronically in ASCII text files. The Bureau is constructing a data dictionary for all approved variables. As a result of this effort, managers and staff will now have greater access to services data. Equally important, the database will capture outcomes oriented, performance based data that will be easily accessible for conducting research and policy studies.

Developing highly responsive databases that enable rapid access to data

The Department of Mental Retardation (DMR) has developed a software program for managing client applications and service delivery. While the primary purposes of the Connecticut Automated Mental Retardation System (CAMRIS) are case management and quality assurance, it also serves as a powerful relational database. The MAPPER software package allows DMR staff to rapidly link different datasets and produce tables that aggregate individual characteristics such as age, sex, race, level of disability, services received and medical information, listing each client in the database by a unique 8-digit identifier. As a result, CAMRIS makes it easy to construct inquiries with several qualifiers and to analyze DMR data in many different ways without an excessive amount of programming.

(Promising Practices continued)

In fiscal year 2001, the Department of Labor (DOL) implemented the automated Case Management and Information System (CMIS) to streamline the administration of its Jobs First Employment Services and the federal Welfare to Work programs. Using a unique identifier, data are stored for each participant in real time (online/ computer screen) as case managers, DSS workers and service providers work with the client. Data are instantly accessible to all users of the system. CMIS captures extensive data on participants that can represent barriers to employment, including education, learning disabilities, substance abuse, child care, transportation and health. Demographic and service delivery data are easily extracted from the system to construct reports and to respond to inquiries.

4. **Train agency staff so that they understand the importance of expanding the functionality of current databases and develop the skills necessary to create databases that support policy-relevant research.**
   To make the transition from an environment that is rich in data but poor in information is a large, but not insurmountable step. One key element is training for state agency staff committed to this process. There may be a need to expose agencies to new skills, training and education that move the mindset from day to day operations toward a vision for impacting the future. For some, this would involve introduction to data management and tabulation, software programs, statistical procedures and general research methods. For others, it could involve an understanding of how evidence can inform policy making. Workshops and consultation sessions could assist both information technology and service managers in establishing a workplace that values and produces policy-relevant information.

5. **Automate data collection where paper filing is currently in use.**
   Valuable data is virtually inaccessible for research and analysis purposes if it is not in electronic form. A top priority should be automating all databases.

6. **Ensure that all data is archived.**
   Without an archive of past data, one cannot assess changes over time, which is a critical component of program and policy evaluation. One is also prevented from matching cohorts of data across databases. To enable these comparative and longitudinal analyses, all databases should contain archive functions and data should be saved systematically.

7. **Develop guidelines and standards for constructing and modifying administrative databases to facilitate research and linking data.**
   Guidelines and standards for new databases, and enhancements of existing ones, would make analysis and new research easier and more inviting. Suggestions are offered below.

   a. **Establish a common identification system.**
      Establishing a common record (e.g. client, patient, family or case) identification number is of central importance to developing the most meaningful research potential from administrative data. The use of a common identification system would enable the linking of databases across agency lines and the matching of client records across programs. It would

facilitate longitudinal and retrospective research designs that track cases over time. It would also allow one to examine patterns of how families use various agency programs, whether they transition from one system to another and whether a constellation of services or a single service produces a certain outcome.

Promising Practices: Linking data for more in-depth analysis

Establishing a unique record identification number for use across databases
The State Department of Education (SDE) has pioneered a single system of identifying students for school and department databases. Each child is identified by a unique string containing full name (last, first, middle) and date of birth (03/14/97). A new consolidated database has been installed at SDE that is created by electronic uploads from school districts' records on students three times a year. This database will grow over the years to generate aggregate reports currently submitted by districts. Created on a student level, it will allow SDE to use the same data in multiple reports and has the potential as a tracking system to allow important research and analysis on student progress.

Linking databases across agencies
A joint project of the state Departments of Mental Retardation (DMR) and Education (SDE) is linking data DMR collects on children in the Birth to Three program with data SDE collects on Preschool Special Education. This system will facilitate referrals and tracking of children with special needs, and will foster accountability, reporting and research on the transition of children from one program to another. Ultimately, these children with disabilities who are identified early will be followed into elementary school and outcomes can be examined.

b. Standardize data coding format.

Coding and recording data in a consistent and research-friendly format will facilitate linkages among agencies and programs and better research capacity overall. For example, one suggestion might be to record the year of birth as a <u>separate</u> variable and not part of a string (e.g. 1975, not 4/1/75). This is very important both for calculating age and also for tracking cases. In the interest of sharing data, the ages for young children should be recorded in a standard format in terms of months or part of the year (e.g. 18 months, 1.5 years, or 1 year 6 months). By developing consensus among state agency partners, the project could give guidance to database managers far beyond early childhood interests.

8. Initiate several pilot projects that demonstrate enhancing and linking of databases for policy-relevant research.

A few pilot efforts undertaken by Early Childhood DataCONNections and state agencies are manageable and realistic. As research pilots are pursued, it is likely that specific ways to enhance particular databases will become evident. These projects have great potential for identifying where the addition of a data point to the larger database would make a valuable contribution to its capacity for research, or where a variable's definition could be modified to bring it in line with that of other agencies. Furthermore, monitoring the process in terms of what did and did not work will help guide future efforts. Beyond the technical advances that can result, these pilots can also generate an awareness of the possibilities and appetite for information on policy issues.

9. **Disseminate information to agencies on best practices.**
   Showcasing success stories and promising practices among agency leadership is needed so that replication can occur. Those in the academic and research communities who have experience or expertise in the conversion of administrative databases into datasets that support policy development are assets. They can and should be engaged to assist in making agencies and researchers aware of the opportunities.

10. **Develop early childhood research partnerships between state agencies and institutions of higher learning.**
    Sharing interests, information and resources is a mutually beneficial arrangement for state agencies and researchers. Through such partnerships, state agencies can meet their research and data analysis objectives and researchers can advance knowledge in their chosen fields and in the policy arena. With each collaboration, expertise and capacity for future work grows. In some states, long-lasting policy research centers on children or early childhood issues have emerged from similar partnerships.

11. **Improve overall access to state data.**
    The data that state agencies collect is a valuable resource for a variety of audiences including, but not limited to, researchers, advocates and the general public. While improving the scope and quality of data collected is critical, it is equally important to share the data in a timely manner and in an accessible way. Most people have come to rely on the Internet for information and state agencies should endeavor to share information, whenever possible, through the web. Furthermore, published reports should include substantive information and should be presented in a reader-friendly format.

CONCLUSION

Using administrative databases as tools for policy-relevant research is a practical and seemingly cost-effective way to promote informed decision-making. With enhancements that enable evidence-based research, state administrative databases have great potential for shaping and advancing sound public policy. The partnerships and resources that the Early Childhood DataCONNections project is developing are the foundation for realizing that potential and for achieving the ultimate goal of producing effective services and positive outcomes for young children and their families.

Appendix A: Survey Methodology

Early Childhood DataCONNections commissioned Words & Numbers Research, Inc. to identify databases used in Connecticut state agencies for aggregating data on young children, specifically infants through age eight, and services for these children and their families. Based on this criteria, eight agencies were selected: the State Departments of Children and Families, Education, Labor, Mental Retardation, Public Health, and Social Services as well as Community Colleges (Connecticut Charts-a-Course) and the United Way of Connecticut (Infoline).

To create the inventory, an interview protocol (see Appendix B) was developed to facilitate in-depth onsite interviews with key administrators. The interview protocol, along with an individual confirmation letter, was sent to each interviewee in advance. Overall, thirty-two interviews were conducted (see Appendix C).

After the interviews, the database materials collected during the interview process were reviewed and analyzed. These materials included data collection forms, computer screen prints, codebooks, data dictionaries and reports. The content of each database was reviewed carefully, with particular attention to which variables were used and how they were categorized. Due to the large number of databases and their variability, a database reporting template was developed to simplify the analysis. The template includes the following basic information:

Template Format

Department:

Bureau / Division:

Database title:

Database description:

Data tabulation and management:

Data collection period:

Data collection software:

Reports produced from data:

The final product is profiles of over 80 databases, including descriptions of the variables contained within. These profiles are maintained in DataCONNections' library. For more information on this resource, contact Sue Wilson at info@chdi.org.

Appendix B: Interview Protocol

Client Population

- What population is covered by your database?
- How is a case or record defined?
- What client or program identifier do you use to input data on service recipients?
- When and how are client or program data collected?

Fields/ Variables

- What fields or variables are collected on clients or programs?
    - Demographic (e.g. birth date, gender, race and ethnicity, town of residence)
    - Service/ process (e.g. presenting problem, type of service provided, unit of service)
    - Outcome (e.g. type of placement, achievement of goal, reason for termination)
- How is each of these fields or variables defined?
- To what extent is complete data provided on the fields? Which, if any, fields are less reliable for planning or reporting purposes due to the lack of data provided?
- What criteria are used to indicate a terminated case? How are terminated cases designated in the database?
- Is data entered on applicants who are denied services? Are reasons specified?
- How is history maintained? Is it retrievable?

Reporting

- What reports does your department issue from this database? Which of these reports are internal or external? Are any of them web-based?
- How often are each of these reports issued?
- What, if any, data is collected but not utilized for reporting purposes?
- What is the protocol for creating a customized report from this database?

Interface

- To what extent does this database interface with other state maintained databases?

New System Development

- What, if any, new systems are being planned to capture client or program data?

Please provide copies of the following:

1. Data collection form or sheet, or screen
2. Code book or record layout format
3. An example of each report issued on a regular basis

Appendix C: Interview Participants

Dept. of Children and Family Services
MaryAnn Dayton-Fitzgerald
Lester Horvath
Celeste Jorge

Dept. of Education

Janet Foster
Camille Jackson-Alleyne
Doug Rindone
Roberta Pawloski
Diane Murphy
Judy Carson
Agit Gopalakrisnan

Dept. of Labor
John Ford
Roger Theirren
Adele De Francesca
Tina McQuiggan
Mark D'Avignon
Steve Litke
John DiSette
Mary Ziomek

Dept. of Mental Retardation
Beverly Bellisio
Alice Ridgeway

Dept. of Public Health
Ardell Wilson
Suzanne Blancaflor
Devin Conover
Lloyd Mueller

Dept. of Social Services
Jan Miller
Donald Beltrane
Neil Newman
Tom Horan

CT Infoline
Sherri Sutera
Dawn Grodzki
Tanya Barrett
Chris Fallon
Kareena DuPlessis
Marijane Carey

Community Colleges
Stacy Garnett

FOR ADDITIONAL COPIES OF THIS REPORT, CALL 860-679-1519 OR

visit our web site:
www.chdi.org

The Child Health and
Development Institute
of Connecticut, Inc.

270 Farmington Avenue
Suite 325
Farmington, CT 06032

860.679.1519 office
CHDI@adp.uchc.edu
www.chdi.org