

# Putting Administrative Data to Work

A Toolkit for State Agencies on  
Advancing Data Integration and Data Sharing Efforts  
to Support Sound Policy and Program Development

April 2005



Child Health and Development  
Institute of Connecticut, Inc.

Prepared by:  
Early Childhood DataCONNECTIONS  
An initiative of CHDI

# Acknowledgements

---

This toolkit was authored by **Frances Duran** and **Susan Wilson**, of the Early Childhood Data-CONNECTIONS initiative at the Child Health and Development Institute, and by **David Carroll**, of Words & Numbers Research, Inc. This body of work is based on research and analysis conducted by David Carroll.

DataCONNECTIONS is grateful to those who dedicated time and expertise to the completion of this toolkit, especially members of the Database Guidelines Toolkit Working Group and others who were interviewed as part of the development of this resource.

## **Database Guidelines Toolkit Working Group**

Preston Britner

*University of Connecticut*

David Carroll

*Words and Numbers Research, Inc.*

Marcie Cavacas

*Department of Public Health*

Frances Duran

*Child Health and Development Institute*

Mary Alice Lee

*Connecticut Voices for Children*

Jan Miller

*Department of Social Services*

Diane Murphy

*Department of Education*

Peter Palermino

*Department of Social Services*

Louis Polzella

*Department of Social Services*

Alice E. Ridgway

*Department of Mental Retardation*

Amy Sampson

*Connecticut Association for Human Services*

Joan Twiggs

*Department of Children and Families*

Susan Wilson

*Child Health and Development Institute*

## **Other Partners**

*Department of Children and Families (DCF)*

Louis Buonincontri

Lester Horvath

Michael Schultz

*Department of Education (SDE)*

Robert Lucco

Douglas Rindone

*Department of Public Health (DPH)*

Nadun Athukorala

Jennifer Morin

Lloyd Mueller

Carol Stone

*Department of Social Services (DSS)*

Donald Beltrame

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

For additional copies of this toolkit, call 860-679-1519 or visit [www.chdi.org](http://www.chdi.org)



Child Health and Development  
Institute of Connecticut, Inc.

Dear Colleagues,

We are stepping into an exciting time in state government. A new realization of the power and usefulness of data is emerging in the public policy and human services arenas because of a convergence of our technological and conceptual advances.

Electronic technology, both computer and web-based, is giving us new tools for collecting, merging, analyzing and reporting data. As a result, there is great potential for generating better and more helpful information for those making policy and programming decisions. Both nationally and here in Connecticut, some have already begun to tap this potential.

At the same time, conceptually we are breaking down the walls of specialization and linear thinking in favor of a more integrated approach toward addressing social concerns and, consequently, using data. For example, in child health, a new definition of health is being promoted that extends beyond the focus on disease and injury to a more holistic vision of living actively and coping with environmental stresses. Assessing progress toward this new standard of health requires us to look outside traditional health data sources.

CHDI and our state, academic, and community partners are pleased to promote the building of strong research and data infrastructures within state agencies to capture these technological and conceptual advances and put them to work in Connecticut. We offer this toolkit to that end.

My thanks go to the working group who guided the development of this work and to David Carroll of Words & Numbers Research Inc. for his extensive research on which the toolkit is based. My special appreciation goes to Frances Duran, the primary manager and author of the effort, whose intellect and creativity have brought it to life.

Sincerely,

Susan B. Wilson, Director  
Early Childhood DataCONNECTIONS  
Child Health and Development Institute of Connecticut

### **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 works to assure that children in Connecticut who are disadvantaged will have access to and make use of a comprehensive, effective, community-based health and mental health care system.

*For more information, visit [www.chdi.org](http://www.chdi.org)*

Any portion of this toolkit may be reproduced without prior permission, provided the source is cited as:

Duran, F., Wilson, S. & Carroll, D. (2005). *Putting Administrative Data to Work: A Toolkit for State Agencies on Advancing Data Integration and Data Sharing Efforts to Support Sound Policy and Program Development*. Farmington, CT: Child Health and Development Institute of Connecticut.

# Table of Contents

---

## Introduction

|   |    |
|---|----|
| Purpose of the Toolkit  | 7  |
| Putting Administrative Data to Work: A FLOW CHART                     | 9  |
| Why Pursue Infrastructure Enhancements?                               | 10 |
| How Will a Strong Data and Research Infrastructure Benefit the State? | 10 |
| How to Use the Toolkit  | 15 |

## Implementation Guidelines

|   |    |
|---|----|
| <b>Getting Started</b>  | 17 |
| Step 1: Secure commitment from agency leadership                  | 18 |
| Step 2: Determine enhancements needs                              | 21 |
| Step 3: Evaluate resources available to address enhancement needs | 23 |
| General Project Management/Implementation Guidelines              | 26 |

|   |    |
|---|----|
| <b>Laying a Strong Foundation: <i>Data Quality Enhancements</i></b> | 29 |
| Establish common record identifiers                                 | 30 |
| Establish common variable definitions and coding                    | 32 |
| Implement/improve archiving functions                               | 34 |
| Reduce missing or erroneous data                                    | 36 |

|   |    |
|---|----|
| <b>Making Connections: <i>Data Linking/Integration Enhancements</i></b> | 39 |
| Link individual databases or data sets                                  | 40 |
| Integrate data and make it easily accessible to end users               | 43 |

|   |    |
|---|----|
| <b>Sharing with Others: <i>Data Access Enhancements</i></b> | 55 |
| Establish Institutional Review Boards (IRBs)                | 57 |
| Develop research protocols                                  | 59 |
| Develop user-friendly codebooks and data dictionaries       | 61 |
| Develop formal data sharing agreements                      | 63 |

|  |    |
|--|----|
| <b>Thinking Strategically: <i>Research Capacity Enhancements</i></b> | 67 |
| Integrate evaluation/ data collection into program design            | 68 |
| Develop a research agenda  | 71 |
| Create a research committee  | 74 |
| Develop research partnerships  | 76 |

|                 |    |
|-----------------|----|
| <b>Endnotes</b> | 79 |
|-----------------|----|

## Resources/References

*Mining an Untapped Resource: Transforming State Data into a Powerful Tool for Decision Makers*, a promotional piece by Early Childhood DataCONNECTIONS

Recommendations from *Reshaping Administrative Databases into Tools for Policy-Relevant Research*, a report by Early Childhood DataCONNECTIONS

List of Recommended Resources

National and Local Enhancement Efforts and Key Contacts

Data Integration Case Studies: CHARM (Utah) and MOHSAIC (Missouri)

Data Sharing Agreement Between the Connecticut Departments of Public Health and Social Services

Connecticut Department of Public Health's Work Plan for HIP-Kids  
(Child Health Informatics Profile Database)

Massachusetts' Business Case for MassCARES - An Integrated Information System

*A Better Way to Kill Dragons: Parody on the Experience of Building an Integrated Information System in Missouri* by Garland H. Land

References

# Introduction

---

Information is a powerful lever, particularly in the field of public policy. Critical policy and programming decisions are influenced by the information that legislators and program administrators receive. This is why it is essential that the data informing these decisions be of high quality (i.e. accurate, timely, and responsive to information needs). A lack of quality data may cause valuable programs to lose funding because their efficacy cannot be determined and vital issues to be continually overlooked because statistics fail to capture the scope of the problem.

**State agencies can play a pivotal role in increasing the pool of quality data and ensuring that it is put to good use.** Most state agencies collect a great deal of both qualitative and quantitative data within their administrative databases. These data, although primarily used for administrative functions like case management and reporting, also have the potential to answer a vast array of policy questions – particularly when different databases are linked together. But, before this potential can be realized, numerous obstacles must be overcome, including incompatible database structures, lack of data quality, limited data sharing, and state agencies’ typically low prioritization of expanding data usage for research purposes.

Recognizing the power of administrative data and wanting to help agencies and policy researchers mine this virtually untapped resource, the Early Childhood DataCONNECTIONS project<sup>1</sup> was created in 2002. One of the project’s first tasks was conducting a scan and analysis of state agency databases<sup>2</sup> to determine what enhancements were necessary to maximize the potential of administrative data as a tool for policy and planning.

#### Administrative data:

*Data collected by state agencies and primarily used for record-keeping and case management; monitoring and evaluating program performance; and ensuring agency accountability.*

As a result of this study, DataCONNECTIONS released a July 2003 report entitled, “Reshaping Administrative Databases into Tools for Policy-Relevant Research” (available at [www.chdi.org](http://www.chdi.org)), which outlined general recommendations for optimizing the utility of state agency data. Many of these recommendations involved addressing the compatibility of databases, as well as data quality and accessibility (a complete listing of the recommendations is available in the Resources/References section). This toolkit is part of the next phase of this effort. It provides state agencies with guidelines on how to implement the recommendations offered in the 2003 report and in doing so, enhance their data and research infrastructure (see description on page 8).

#### 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.*

## Purpose of the Toolkit

As stated above, the ultimate goal of this toolkit is to help state agencies strengthen their data and research infrastructure. It provides an assessment tool to help agencies determine their enhancement needs as well as guidelines on how to approach implementation of several different infrastructure-enhancing strategies. The toolkit is not intended to function as a technical design

manual, but rather seeks to help agencies identify necessary components for successful implementation of the strategy or strategies they choose to pursue. The guidance provided within the toolkit is based on best practice and lessons learned from those that have worked on similar efforts, both nationally and in Connecticut. The toolkit also serves as a valuable reference document, providing links or citations for relevant resources, as well as contact information for those wanting to follow up with key personnel involved in many of the efforts highlighted.

In addition, **the toolkit seeks to raise awareness of critical components that underlie the development of a strong data and research infrastructure, and to guide efforts toward putting these building blocks into place.**

These key components include:

- *Committed agency leaders* who recognize the value of high quality data and elevate its importance within the agency
- *A sound plan* for developing, implementing and sustaining a strong infrastructure
- *Adequate technology* that maximizes data utility and accessibility
- *Accurate data entry* that reduces inefficiencies and erroneous information
- *Data collection* that is sufficient to address administrative needs as well as key research/planning questions
- *Staff resources* for data entry, training, information management and data analysis
- *Protocols and partnerships* that facilitate access to agency data for higher education and other research institutions

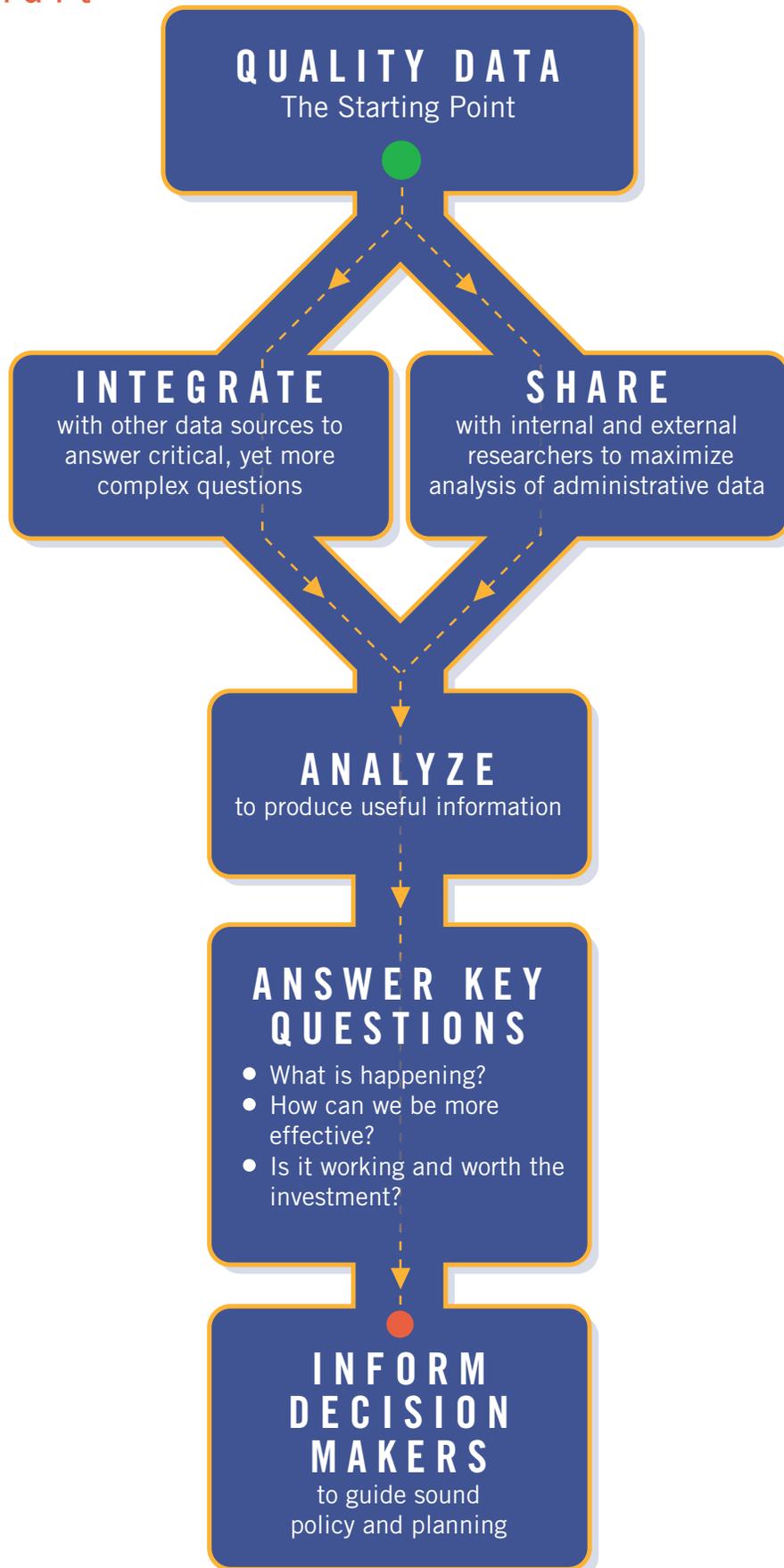
### What Does a Strong Data and Research Infrastructure Look Like?

- Data is of high quality (accurate and aligned with programming, planning, and policy information needs)
- Databases have the capacity to link to other databases, within and outside of the agency
- Data archiving and/or storage procedures are in place to enable longitudinal analysis
- “Public use” data is available to the public in a timely manner and user-friendly format

### NOTE:

The flow chart on the following page further illustrates the infrastructure envisioned – quality data is integrated, shared, and analyzed to produce reliable and useful information that will inform sound policy and planning. Maintaining data quality, of course, should be an ongoing effort.

## A Flow Chart



## Why Pursue Infrastructure Enhancements?

Many state agencies tend to be data rich, but information poor. This paradox largely stems from low prioritization of enhancing information systems and promoting the use of administrative data for research. Typically, state agencies' limited resources are channeled more directly toward client service delivery, which is their highest priority. As a result, agency data management systems are often fragmented and provide agencies and researchers with little flexibility in analyzing the data. Furthermore, few agencies promote the use of administrative data for research purposes through activities like setting a research agenda, institutionalizing data access protocols, or establishing public/private research partnerships.

However, efforts are underway to help agencies envision and pursue data and research infrastructure enhancements as a strategy for improving client service delivery and overall child and family outcomes. On the national level, the Robert Wood Johnson Foundation is funding *All Kids Count*, an initiative that is designed to foster development of integrated child health systems within state public health agencies in order to improve children's health.<sup>3</sup> In addition, a recent report by the Committee on Evaluation of Children's Health, which was formed by the National Research Council and the Institute of Medicine, identified the need for improved data collection, data linkages across various data sources, and improved access to health information by the research communities as critical to helping federal, state and local entities accurately assess and promote children's health.<sup>4</sup>

Data linkage: *The process of combining two or more unique databases (or datasets) in an ad hoc manner into a meta-database by matching unique identifiers within the databases.*

These efforts and many others highlighted throughout this toolkit underscore the value of data and research infrastructure enhancements as strategic investments for agency leaders and all others interested in the reliability and substance of information provided through administrative data. For example, enhancing data quality yields a number of cost savings, including greater efficiency when conducting state research and greater accuracy in determining who is (and is not) eligible for services. Furthermore, initiatives that improve coordination and integration across information systems not only provide more in-depth information on children and families, but also reduce agency costs by avoiding the duplication of effort that comes with maintaining individual data management systems. These integration efforts benefit agency clients as well by alleviating the burden that comes with having to interface with multiple information/intake systems.

## How Will a Strong Data and Research Infrastructure Benefit the State?

Overall, a solid data and research infrastructure can help further many agency goals and objectives and benefit a variety of stakeholders. The following list highlights some of the benefits ensuing from various components of a strong infrastructure and includes specific examples of how these benefits have materialized in different states.

## *A Strong Data and Research Infrastructure Helps to...*

### **1. Contain programming costs and reduce fraud.**

Data that can accurately evaluate service delivery and track individuals across time and programs can help save program administrators money by improving service delivery efficiency and zeroing in on fraudulent activity. In addition, cost-savings accrue when data systems are more compatible and linking or extracting data becomes less cumbersome.

***Example:** Michigan Department of Community Health (MDCH) Enterprise Data Warehouse*  
Through implementation of the data warehouse and, consequently, higher quality data, MDCH was able to conduct more effective surveillance of fraud and abuse and improve cost containment.

### **2. Expedite and improve reporting to federal government, policymakers and others.**

With better quality data – stored in databases (or data warehouses) with enhanced data retrieval mechanisms – not only is the information being reported more accurate or relevant, but it is easier to access.

***Example:** Wisconsin Data for Organizational Management (WISDOM)*  
Wisconsin's integrated reporting and query system provides easy access to data on social service programs, including TANF, Food Stamps and various medical assistance and child care programs. WISDOM allows users to conduct their own queries and reports without involving information technology (IT) staff. The user interface makes it easy for agency staff to generate reports internally and for external audiences such as legislators and the media.

### **3. Improve public access to information.**

Having data sharing protocols in place that protect client confidentiality and creating tools like data dictionaries that describe agency data in detail significantly improves access to agency data. Access for the general public is also greatly enhanced through agencies' timely, online posting of statistics and research findings to a website with a user-friendly interface.

***Example:** North Carolina State Center for Health Statistics (SCHS)*  
SCHS places data from health surveys and registries on their state department website for easy user access. The public can search the site ([www.schs.state.nc.us/SCHS](http://www.schs.state.nc.us/SCHS)) for information by topic (e.g. birth defects, vital statistics) as well as by publication title. The site is continuously updated, providing users rapid access to the most recent data and studies.

### **4. Protect client confidentiality.**

A strong data infrastructure might include several different elements that protect confidentiality, such as an Institutional Review Board with oversight as to how client-level data can be used

***Data warehouse:** A repository for data integrated from multiple data sources that is centrally located, automatically fed regular data updates, and that generates data extracts for end users.*

and/or a data warehouse designed to automatically restrict access to certain data based on a requestor's clearance level.

***Example:*** *Washington State Institutional Review Board*

Washington State has a joint Institutional Review Board that reviews and approves all research sponsored by the Department of Social and Health Services (DSHS), the Department of Health (DOH), or the Department of Labor and Industries (L&I), as well as research conducted by one of these state agencies or using any non-public information held by these state agencies. The review process is intended to protect the rights and welfare of subjects participating in the research, and to assure that the research is sound and produces benefits that outweigh the risks to the subjects. The review also protects the departments from liability resulting from improperly conducted research. See [www.spokane.wsu.edu/research&service/HREC/IRB](http://www.spokane.wsu.edu/research&service/HREC/IRB)

#### **5. Support sound policy development and planning**

Quality data that is timely, accessible and significantly informative can yield enormous benefits to those in charge of policy development and planning. The ability to clearly and accurately look at what is happening within and across programs/service systems, particularly from a longitudinal perspective, is a powerful tool. Several examples are provided below that highlight the various functions that high quality data can help state agencies and others to perform. Pursuing technical upgrades and engaging researchers in strategic discussions about data collection all contribute to the proliferation of quality data and the establishment of a strong data and research infrastructure within an agency.

***Examples:*** *See supplement on the following page, "Strengthening Policy Development and Planning through Data."*

#### **NOTE:**

---

Additional information on many of the examples provided above or throughout the toolkit may be found in the Resources/References section. For those initiatives not catalogued in the back of the binder, contact information and/or web addresses are provided with the example in the body of the toolkit.

---

## Strengthening Policy Development and Planning through Data: Examples from Across the Country

### **Identifying Areas of Need**

*Childhood Asthma in North Carolina (State Center for Health Statistics)*

This study examined the prevalence of asthma among children on Medicaid and the rates of hospitalization for asthma among all children in the state. Two sources of data were used to estimate the prevalence of asthma among children in North Carolina: Medicaid paid claims and hospital discharges.

*The Use of Public Health Databases to Estimate the Risk for Special Education Placement (State Center for Health Statistics)*

This study examined the risk factors associated with placement in special education in the first grade. The purpose was to help identify children in need of early intervention services. Records for children enrolled in first grade special education were linked to birth certificate records. Identification and referral records from Child Service Coordination (CDC), an early intervention and service coordination program for young children, were also linked to identify children who received CDC services and the associated risk conditions.

### **Assessing Service Utilization**

*Lead Screening Coverage for North Carolina's Medicaid Children (State Center for Health Statistics)*

The purpose of this study was to provide a complete count of Medicaid children who had an opportunity to receive a blood lead test through a Health Check (well-child care) visit in 1998 and 1999 and to calculate the percentage of those children who received a blood lead test in 1998 and 1999. Medicaid claim records were matched to lead test laboratory records.

*Child Care Subsidy Dynamics Study Team*

A consortium of universities used state administrative data from five states (Oregon, Texas, Illinois, Massachusetts, and Maryland) to study their child care subsidy programs in order to better understand who is served by the programs, what services they receive, and how long they are served. See <http://www.hhs.oregonstate.edu/familypolicy.occrp/>

Contact: Roberta Weber, Oregon State University, [bobbie.weber@oregonstate.edu](mailto:bobbie.weber@oregonstate.edu)

### **Measuring Program Efficacy**

*Prenatal WIC Participation in Relation to Low Birth Weight and Medicaid Infant Costs in North Carolina (State Center for Health Statistics)*

The purpose of this study was to determine the extent to which prenatal WIC participation resulted in improved birth outcomes and reduced infant medical costs. Medicaid and WIC data files were linked to 1997 North Carolina live birth records.

*The Patterns of Food Stamp and WIC Participation and their Effects on the Health of Low-Income Children (Illinois Integrated Database on Child and Family Services, Chapin Hall)*

This study examined the patterns of program participation in Food Stamps (FSP) and WIC, and the effects of WIC on young children's health outcomes. The study used a linked data set based on population-level administrative data on all births, Food Stamp and WIC participation, and Medicaid eligibility and claims in Illinois between 1990 and 1998.

---

## Strengthening Policy Development and Planning through Data: Examples from Across the Country (continued)

### Improving Program Efficiency

*Trends in Emergency Department Utilization (RIte Care Medicaid Managed Care Program)*

This study analyzed emergency department utilization rates by quarter and treatment category, as well as by age, gender, diagnosis and average cost of visit by treatment type. The study also examined the extent to which enrollees were being channeled to the most timely, clinically appropriate and cost effective treatment options. Quarterly encounter data were used for this evaluation.

### Measuring Effects of Policy Changes

*Infant Mortality and Low Birth Weight in North Carolina (State Center for Health Statistics)*

This study used matched birth and infant death certificates to examine changes in the birth weight distribution as well as birth weight and cause-specific infant mortality. The purpose was to discern trends and patterns in the ten years since high infant mortality rates in North Carolina resulted in expanded maternal and child care programs, including the expansion of Medicaid coverage for pregnant women.

*Changes in Child Social Program Participation in the 1990's (Illinois Integrated Database on Child and Family Services, Chapin Hall)*

This study examined the effects of social policy changes on low-income children and families by monitoring changes in the utilization of social programs among children who entered AFDC/ TANF during the 1990's. It used individual-level administrative data to follow children who entered AFDC/ TANF for the first time in the 1990's and to describe their foster care placement, and abuse and neglect report experiences. Their patterns of exit from AFDC/ TANF into Medicaid and Food Stamps were also examined.

*Rhode Island Child Care Research Partnership (Wellesley College)*

Rhode Island, through Wellesley College, studied the child care choices of 19,386 Rhode Island families that received child care subsidies over a four year period. Using state administrative data they were able to examine the impacts of the state's extensive child care reforms on the actions of these families and differentiate by household and community characteristics.

Contact: Ann Witte, Wellesley College, [awitte@wellesley.edu](mailto:awitte@wellesley.edu)

---

As the preceding examples illustrate, quality data can greatly assist state agencies in their daily work. **These examples also highlight how a database's utility is exponentially expanded if it has the ability to link to other databases.** Linked data may be the product of simply combining two or more databases in an ad hoc manner or a more complex and enduring effort, such as creating a data warehouse. Either way, for many agencies, enhancements are needed to successfully and efficiently link data. Many of the enhancement strategies discussed in this toolkit are focused on supporting data integration.

**Data integration:** *The process of merging and consolidating data from a variety of sources into a single data source or platform, such as a data warehouse.*

## How to Use the Toolkit

Although the toolkit is geared toward state agency leadership and staff, it will likely also be of interest to others (e.g. researchers) who work closely with administrative data and have a stake in improving data quality and access.

In using the toolkit, it is important to note the following:

- The toolkit is not a technical design manual, but rather *a guidebook and reference manual* for those interested in pursuing enhancements and/or learning about enhancement initiatives taking place in Connecticut and other states.
- The toolkit *highlights key enhancement strategies and offers guidance* on critical steps toward pursuing those enhancements. Neither the strategies nor the guidelines should be viewed as exhaustive.
- The information presented is *intended for a wide range of agency audiences, with varying levels of expertise* in and understanding of data management and research design. Hopefully, the toolkit format will make it easy for the various users to find the information that is most relevant to them.

It is also important for readers to know that all involved in the toolkit's development recognize that Rome was not built in a day. Achieving a strong data and research infrastructure that promotes the use of administrative data for research purposes and yields accessible, high-quality data will surely be a gradual process. One of the toolkit's primary goals is simply to help this process move *forward*.

State agencies' priorities, opportunities and resources will have a large impact on which and how many enhancements the organization is able to pursue. Therefore, this toolkit is designed with an incremental approach in mind. **Each strategy that is outlined in this toolkit is individually significant, yet also represents a building block toward establishing a strong data and research infrastructure and ultimately, optimizing the quality, utility and accessibility of state data.**

The toolkit outlines a spectrum of different enhancement strategies. A good number of the strategies addressed center around data quality – improving data accuracy as well as the scope of data collected. Many are also technical upgrades that will assist in linking databases and provide a good foundation for data integration. Other enhancements are more focused on making sure data is accessible so that it can be analyzed and provide useful information to agency staff, policymakers and other key stakeholders. This is a critical piece of a strong data and research infrastructure. State agencies often do not have the resources to analyze their data and many researchers are eager to access the wealth of data housed in state agencies, so fostering relationships with researchers external to an agency is a mutually beneficial activity. Finally, the last grouping of enhancement strategies describes ways to entrench data and research within agency culture through activities such as integrating evaluation into program design or establishing research partnerships with higher education institutions.

It is the vision of the team that developed this toolkit, that all state agencies serving children and families will be able to creatively and cooperatively develop a culture and infrastructure that values and supports research and ultimately evidence-based decision making on public policy issues. There is no doubt that such a vision can be achieved, but it requires partnerships across agencies and with the research community.

In the next section, the toolkit walks users through the process of determining what enhancements need to be made and making decisions about which to pursue. The following sections provide guidelines on how to implement the various enhancement strategies outlined above and, in many instances, provide examples of other states/organizations that have undertaken similar efforts.

Additional information on many of these efforts, as well as additional resources for those pursuing enhancements, is also available in the Resources/References section.

## Toolkit Enhancement Strategies

### *Laying a Strong Foundation: Data Quality Enhancements*

- Establish common record identifiers
- Establish common variable definitions and coding
- Implement/improve archiving functions
- Reduce entry of erroneous data

### *Making Connections: Data Utility Enhancements*

- Link individual databases or data sets
- Integrate data and make it easily accessible to end users

### *Sharing with Others: Data Access Enhancements*

- Establish Institutional Review Boards
- Develop research protocols
- Develop user-friendly codebooks and data dictionaries
- Develop formal data sharing agreements

### *Thinking Strategically: Research Capacity Enhancements*

- Integrate evaluation/data collection into program design
- Develop a research agenda
- Create a research committee
- Develop research partnerships

It is the vision of the team that developed this toolkit that all state agencies serving children and families will be able to creatively and cooperatively develop a culture and infrastructure that values and supports research and ultimately, evidence-based decision making on public policy issues.

# Getting Started

---

Hats off to you and your agency for taking – or at least considering – steps toward strengthening administrative data to better serve as a tool for sound policy and program planning. It is a wise investment. This section outlines some critical elements that will help lay the foundation for your enhancement efforts. These include:

**Step 1:** Secure commitment from agency leadership.

**Step 2:** Determine enhancement needs.

**Step 3:** Evaluate available resources and prioritize efforts.

At the end of the section, there are also some general project management/implementation guidelines that are applicable to many of the data enhancement strategies outlined later in the toolkit.



## STEP 1

### Secure commitment from agency leadership.

Without support, particularly from commissioners and other agency leadership, making forward progress or building data and research infrastructure is unlikely. The key ingredient of all successful data enhancement projects is the realization of top leadership that data, research and analysis will enhance the service delivery of their departments. Executive leadership is crucial throughout the process because of the resource allocation involved. Planning, design, implementation and administration of improvement efforts are often time consuming and require many staff and, sometimes, outside experts as well. Depending on the scope of the project, substantial cost may also be involved. Only the commitment of the organization's top executives can provide project planners with the resources and political support they need to overcome any obstacles that they may encounter.

To help a project succeed, the Public Health Informatics Institute (PHII), an organization studying best practices in information systems integration, asserts that two levels of leadership support are required: an executive sponsor and a champion. In some cases, these may be the same person. As explained in a recent PHII planning and development manual, "the executive sponsor is a high level official who works for the institutionalization of the project, creates a work environment that fosters risk-taking and innovation, is a good communicator, and has political awareness and influential contacts. The champion has a passion for the project, the respect of other staff and higher-ups, access to senior leadership, and is willing to devote effort to see the project succeed." (*Integration Sourcebook*, p. 10). If the champion and executive sponsor are not one and the same, it is the champion who is often leading the campaign to engage an executive sponsor. Cultivating this sponsorship may be the hardest thing to do, particularly since the sponsor may change over time given changes in state administration. Below are some tips to help champions and other committed stakeholders with this challenging, but critical task.

### Tips for Cultivating and Maintaining Leadership Support

This toolkit in and of itself may help garner support. For example, those advocating for enhancements can draw upon the various benefits of data quality that are outlined in this document and supported by "success stories" from across the country. The following example from Rhode Island may also be useful to those thinking strategically about methods to engage agency leadership in data improvement efforts.

***Example:*** *Rhode Island's Children's Cabinet*

In order to demonstrate the feasibility and practicality of sharing data among several state child-serving agencies and to obtain commitment from the leadership of the various agencies, the Cabinet decided to attempt record matches. This exercise not only revealed how many clients the

A great tool to help generate interest in data and research enhancement efforts is:

***Mining An Untapped Resource:  
Transforming State Data into a  
Powerful Tool for Decision Makers,***

a colorful pamphlet for state leaders stressing the value of data in policy and program planning, and highlighting success stories from Connecticut and across the country. Available in the Toolkit's Resources/References section and at [www.chdi.org](http://www.chdi.org).

agencies held in common among their caseloads, but it facilitated analysis of which data fields could be used to establish a common identifier system, using either a new case number assignment or an algorithm for record matching. In the end, the record-matching demonstration convinced agency leadership of the opportunities inherent in data integration (e.g. coordinated service delivery) and the need to put resources into the effort. Contact: Sherry Campanelli, Rhode Island Department of Human Services, [scampane@gw.dhs.state.ri.us](mailto:scampane@gw.dhs.state.ri.us)

Some other suggestions for enlisting leadership commitment are listed below.

- Develop a strong proposal to present to agency leadership that outlines what needs to be addressed, the advantages of pursuing enhancements, and (possibly) the project costs. Make sure the proposal clearly articulates the *specific* benefits to the agency. The proposal might recommend hiring a consultant to develop a strategic plan and provide cost projections.

- Arrange for a meeting between agency leadership and another top-level person from an agency or organization that has successfully implemented data enhancements and can attest to the value of such efforts. Individuals from out-of-state or a national organization may be best to engage in this conversation.

**Example:** *Linking Health Data Within the Connecticut Department of Public Health (DPH): A Roundtable Discussion*

To inform development of their Child Health Informatics Profile, DPH convened a roundtable featuring speakers from Rhode Island and Michigan who discussed their successful data integration efforts and shared key lessons learned.

### **Policy Reports that Demonstrate the Potential of Connecticut's Administrative Data**

#### ***A Research Perspective on the Child Care Workforce in Connecticut***

This report summarizes Connecticut data and relevant national information on the child care workforce, including state Head Start and School Readiness employees. It highlights the qualifications, compensation, turnover and number of staff who teach and care for young children in the state, as these factors are known to be predictors of child care quality and, ultimately, to affect child outcomes. ([www.chdi.org](http://www.chdi.org))

#### ***Births to Mothers in HUSKY A***

By linking Connecticut's 2000 birth registry data with HUSKY A enrollment and encounter data, the Children's Health Council was able to conduct valuable policy-relevant research, which was reported in this publication. The study examined the prevalence of prenatal care services provided to HUSKY-enrolled mothers and subsequent birth outcomes, and drew comparisons between mothers who were and were not enrolled in HUSKY A during pregnancy. It also examined health care for enrolled mothers and infants in the first year of life and the association between poor birth outcomes and future health problems. ([www.ctkidslink.org](http://www.ctkidslink.org))

#### ***Keeping Children on the Path to School Success: How is Connecticut Doing?***

Using administrative data from various state agencies, this report examines how Connecticut's young children (birth to age 6) are progressing toward school success by looking at key factors influencing child health and development. The report is designed to help track these factors over time and measure the impact of policy changes as well as the need for additional services. ([www.chdi.org](http://www.chdi.org))

- Convene agency leaders to discuss the potential benefits of pursuing enhancement efforts. To underscore the value of integrating and/or linking data, part of this discussion should focus on overlaps in clients served across programs and agencies.
- Demonstrate the potential of administrative data to inform planning and policy by conducting analyses and publishing reports (see sidebar on preceding page).

## STEP 2

### Determine enhancement needs.

Creating a strong data and research infrastructure is no easy task. There are many steps in this process, spanning from improving data quality to integrating data to sharing data internally and externally, especially with researchers. Therefore, it is important to understand at the beginning what needs to get done in the broadest sense and then, one can develop a plan for addressing these needs at a pace that is aligned with agencies' resources and priorities. However, assessment should be an ongoing process to ensure that a strong infrastructure is ultimately achieved and maintained. The following checklist can help agencies in this process.

#### ✓ Data Research and Infrastructure Assessment Checklist

##### **Attitudes Toward Data and Research**

- ✓ Agency leadership views quality data as a critical tool in agency operations.
- ✓ Agency staff understand how quality data is useful in their work.
- ✓ Agency leadership values and uses research in the planning and decision making process.

##### **Utility of Existing Data/Data Systems**

- ✓ Agency staff is able to respond to requests for information from legislators, agency leadership and others efficiently and with confidence in the information being provided.
- ✓ Existing data sufficiently address key policy questions.
- ✓ Existing data systems adequately perform the administrative functions needed by the agency.
- ✓ Existing systems efficiently produce required federal and state reports.

##### **Database Compatibility (Inter- and Intra-Agency)**

- ✓ All databases within the agency use the same common record identifiers.
- ✓ The common record identifiers used by the agency are compatible with those in other state agencies or there is sufficient commonality in the variables collected to facilitate accurate record matching across agencies.

#### **Additional Assessment Resource**

##### *Tool for Assessment and Planning* by All Kids Count

A companion to *Integration of Newborn Screening and Genetic Services Systems with Other Maternal & Child Health Systems: A Sourcebook for Planning and Development*. Although developed specifically to address integration of child health information systems, this assessment tool is useful in the planning and implementation of any information systems project.

Available at  
[www.phii.org/IntegrationSourcebook.html](http://www.phii.org/IntegrationSourcebook.html)

- ✓ All databases within the agency use common variable definitions and coding.
- ✓ The variable definitions and coding schemes used by the agency align with those in other state agencies.
- ✓ Linking and/or interfacing with other agencies' databases is efficient, from both a financial and technological perspective.

### **Variable Coding for Race and Ethnicity**

The US Census has developed guidelines on coding race and ethnicity based on significant changes recommended by the Office of Management and Budget (OMB) regarding the collection of race data for federal agencies. For more information, visit [www.census.gov/population/www/socdemo/race/racefactcb.html](http://www.census.gov/population/www/socdemo/race/racefactcb.html)

#### **Quality Control Measures**

- ✓ The data collected in agency databases is robust enough to answer key research and planning questions, as well as be responsive to standard reporting requirements.
- ✓ All agency databases include features that help prevent data entrants from skipping data fields or entering “dummy” values.
- ✓ All agency databases include archiving functions or, processes are in place to routinely archive data.
- ✓ All staff performing data entry have been trained and educated on proper procedures and are able to follow these procedures without undue burden.

#### **Data Access**

- ✓ Extracting data from agency databases is not difficult.
- ✓ Public use data and reports are accessible online or, are available in a user-friendly electronic format.
- ✓ Data is available to policymakers and the general public in a timely manner.
- ✓ Codebooks that clearly define the variables included in agency databases are available to users both inside and outside of the agency.

#### **Agency Processes and Organization**

- ✓ Agency has a structure and process for evaluating and responding to data requests from other agencies and researchers.
- ✓ Agency has a proactive group that plans for research and data analysis.

## STEP 3

### Evaluate available resources and prioritize efforts.

Once the overall areas for improvement are identified, the next step is determining what is feasible for the agency. It is unlikely that a state agency will have the resources to pursue all the necessary enhancements at once, or even in close succession. To determine what can be done at any given time, agencies must do a crosswalk between resource evaluation and prioritization. However, current projects should be embedded in a long-term strategic plan for strengthening the agency's data and research infrastructure so there is constant forward motion.

### Projects should be embedded in a long-term strategic plan for strengthening the agency's data and research infrastructure so there is constant forward motion.

Depending on how many improvement areas are identified through the assessment process, it may be necessary for agencies to narrow down the number of enhancements being considered. This process should be informed by agency goals and priorities. Some important questions for agencies to consider are:

- What are the agency goals and how can data support these goals?
- Ultimately, what does the agency want to get out of its data management systems?

However, determining what enhancements to pursue should also be guided by the principle that quality data is the foundation for good information. Integrating or sharing data of poor quality does little to support sound decision making at the programmatic or policy level. On the contrary, it can lead to misguided policies and service delivery. As discussed later in this toolkit, there are several components to quality data, including accuracy and consistency in data collection methods. Working to strengthen these components should be given high prioritization.

Once agencies have narrowed down the enhancement projects for consideration, it is necessary to evaluate resources. One method that is often used to facilitate this process is conducting a **feasibility study**. This is particularly valuable when contemplating larger efforts, such as integrating data from multiple sources, but is also useful for smaller-scale projects. In addition to helping agencies assess whether or not the necessary resources are available to pursue the identified improvements, a feasibility study can help inform a cost-benefit analysis. It can also help determine whether or not the enhancement(s) being considered will provide the functionality envisioned by agency leadership, the project planners and end-users. The process of conducting a study also provides an opportunity to identify potential challenges and examine different project design alternatives. Ultimately the feasibility study is a tool for exploring possibilities and informing a reasonable and well-designed implementation plan.

Generally, states have broken their feasibility studies into three broad issue areas: technical, programmatic, and attitudinal. A general framework for a feasibility study, which includes these three areas of consideration, can be found below.

---

## Feasibility Study Framework

### Technical Questions for Consideration

- ? Does the agency have the expertise to design and implement this enhancement internally or will the agency need to hire a consultant for some or all of the work?
- ? Will the agency need to purchase new software or hardware to successfully achieve this enhancement? If so, will the agency opt to purchase *packaged* products or work with a vendor to create something *customized*?

**Note:** Both options have pros and cons. Packaged products tend to be less expensive, but they are also less flexible and may not be able to perform all the functions the agency wants or needs from them. Conversely, customized products are often more expensive, but can be tailored to meet each agency's specific needs.

**Example:** *Child Care Management Information System (CT Department of Social Services)*

DSS faced numerous challenges when trying to adapt a vendor's existing child care subsidy data management system to meet its specific business needs. Among the many obstacles, foremost was the amount of responsibility placed on the end user to develop numerous "work-arounds" to deal with the system's pre-existing structure. In addition, field workers reported that data entry screens did not follow a logical flow and some data elements could not be captured given the system structure. However, by using an existing system over a customized system, it took less development time resulting in a smaller price tag and an earlier launch date. Plus, concerns about operational problems once the system "went live" were minimized because, unlike a customized system, an existing system has been pre-tested.

### Programmatic Questions for Consideration

- ? What are the expectations of the end product's purpose and capacity by planners? Users? Other stakeholders?
- ? How does the cost of each alternative being considered measure against the potential commitment of financial resources?
- ? What time commitments from agency staff and others are necessary for development and implementation of each of the alternatives being examined?
- ? What training will be required once enhancements are in place? Who will develop this training? How many people will need to attend? What will the length of the training be? Will there be need for ongoing technical assistance?

### Attitudinal Questions for Consideration

- ? Is there staff bias against or allegiance to certain hardware, software or platforms?
  - ? Are you likely to confront resistance to implementation from state or local users?
  - ? What constituent concerns might you encounter regarding data access (e.g. confidentiality concerns)?
- 

After assessing resources and determining which enhancement(s) are feasible, it will likely be necessary to again go through a process of prioritization to determine the scope of the enhancement initiative. For example, if an agency decides to begin work on standardizing variable coding and definitions, it may be too overwhelming to commit to addressing every variable in every agency database at once. Instead, the agency may want to develop a work plan that will phase in this enhancement activity, working on batches of variables at a time in their order of priority within the agency. It is a good idea to engage researchers and analysts in the prioritization process so they can help agencies determine what enhancements, at a minimum, are necessary to begin building an infrastructure that will support strong, credible research that is relevant and meaningful to the agencies.

Engage researchers and analysts in the prioritization process so they can help agencies determine what enhancements are necessary to begin building an infrastructure that will support strong, credible research that is relevant and meaningful to the agencies.

## General Project Management/Implementation Guidelines

The following general guidelines for project management are applicable to many of the enhancement strategies outlined in this toolkit, particularly those enhancements focused on data quality and data linkage/integration. More specific guidance relevant to individual strategies is outlined in the next few sections.

### **Develop a strategic plan.**

A good strategic plan specifies the issues that the project is intended to address, the goals it should achieve, the obstacles it will encounter and the key individuals and groups to be involved. Identifying project goals is one of the most important aspects of the strategic plan. Goals are based on the identified business needs and spell out what outcomes are expected. When developing a strategic plan, it is important to think expansively and consider all the functions that agency data should perform, both now and in the future. Then, work on incorporating this vision into the strategic plan and ultimately the project design. Without thoughtful, long-range consideration of what functions the end product should perform, key data may be omitted or captured in such a way that their utility is compromised.

***Example:** Missouri Health Strategic Architectures and Information Cooperative (MOHSAIC)*  
From the outset, executive leaders stated that MOHSAIC data would be used not just for programmatic reporting but also for improving service delivery. As project implementation unfolded, however, planners began to realize that program accountability had been left out of the equation. Funding streams were not being captured and categorical accountability suffered. As a result, program managers were not held fully accountable for the number of services and the outcomes of the services rendered from the resources provided to the program.

### **Establish a steering committee to oversee project planning and implementation**

To ensure that the project maintains its momentum and focus, a steering committee led by an overall project coordinator should be formed. The size and expertise of the committee will be determined by the scope of the project to be undertaken. Committee composition might include senior agency leadership, key program administrators and analysts, information technology (IT) staff and consultants. Regardless of the committee size or composition, the first order of business should be clearly establishing the project objectives and making sure they align with the organization's needs for information as well as functionality. For example, if the committee were developing a data warehouse, they might establish the following objectives:

- Create linkages between the agency's data sources
- Improve ability to respond to query-based data requests
- Improve ability to meet various reporting requirements
- Develop standards that govern the collection and usage of data, agency-wide
- Develop the capacity for tracking clients across agency programs
- Improve capacity to evaluate the efficacy of services provided and to address identified key policy questions

Clearly, there may be many incremental steps needed to achieve the stated objectives. To help ensure that the project stays on task and on time, the committee should also develop a detailed implementation plan and monitor the execution of that plan. The implementation plan should outline reasonable timelines for completion of each task as well as who is responsible for each task and what resources are required.

### *Steering Committee Responsibilities*

- Establish clear project objectives.
- Develop an implementation plan.
- Establish a process for the discussion and resolution of problems.
- Keep members of the organization informed about project activities.
- Cultivate staff ownership of the project throughout the agency.

### **Develop support for data improvements agency wide**

Data quality impacts and is dependent upon the whole organization, from those in the information technology (IT) department to those on the front lines administering programs. Therefore, as planning and implementation efforts move forward, it is crucial to have the support and input of a variety of data users – not just those on the steering committee. Surveying and addressing the needs of many different data users will not only strengthen the end product, but will increase the likelihood that staff will effectively utilize the new and improved data system, data entry procedures, etc.

### *Who Needs To Be On Board?*

- Commissioners
- Program administrators
- Data managers
- Information technology staff
- Caseworkers/Data entry staff

Below are several suggestions for bringing and keeping key stakeholders on board.

- Engage key stakeholders throughout the process. Ensure that enhancements reflect the needs of those who use the data and support program operations.
- When implementing new data systems or processes, make them easy to use and invest in adequate training (more detail on training in the next section).
- Prepare and present materials that are tailored to the audience, answer their key concerns and give the level of detail necessary to their understanding at the particular stage of the process. Examples and graphic representations are usually helpful. A demonstration of the proposed data model and its capabilities can be particularly energizing.
- Use presenters or spokespeople who are valued and trusted by the audience. Try to energize staff through demonstrations at the front end of the project that show the benefits of the new software, data system structure, etc.

In Michigan, the Department of Community Health launched a **marketing campaign** to inform users of the benefits and new functions associated with their data enhancements.



## Laying a Strong Foundation: *Data Quality Enhancements*

---

The cornerstone of a strong data and research infrastructure is high-quality data. Quality not only encompasses accuracy, but also the data's alignment with program administrators' and policy makers' key information needs. Achieving this alignment requires thoughtful strategic planning around data collection and management.

There are many different ways to enhance data quality and this section will address four major enhancement strategies:

- Establish common record identifiers
- Establish common variable definitions and coding
- Implement/improve archiving functions
- Reduce missing or erroneous data



# ENHANCEMENT STRATEGY 1

## Establish common record identifiers.

### Overview

A record identifier is used to identify a client, patient, family or case within a database. Unfortunately, databases within and across agencies develop identifiers in different ways (e.g. some use Social Security numbers while others use computer generated numbers). Therefore, one client served by multiple programs or state agencies can have multiple identification numbers. This can lead to duplication and inaccuracies when tallying numbers served, etc.

Establishing a common record identification number within and across agencies is of central importance to developing the most meaningful research and analysis potential from administrative data. The use of a common record identification methodology enables the linking of databases across agency lines and the matching of client records across programs. It facilitates longitudinal and retrospective research designs that track cases over time. It also allows 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.

Helping to secure privacy and confidentiality is another benefit of creating unique record identifiers. Once a unique identifier has been appended to an individual record, personally identifiable information, such as name, can be secured at the highest level. Individual non-identifying data can then be accessed using identifiers and made available for research purposes.

### Guidelines for Successful Implementation

1. *Inventory common data fields in databases across agencies or units.*
2. If there is not a single, unique field (e.g. Social Security Number, Medicaid case number) that is common to all databases, consider one of the following options:
  - a. *Create a unique computer-generated identification number for each record in a standardized way in all databases to identify the same individual across databases.*
  - b. *Assign a unique record identifier after conducting a record-matching process to eliminate duplicate records.*

Establishing a common record identification number within and across agencies is of central importance to developing the most meaningful research and analysis potential from administrative data.

## A Word About Probabilistic Matching

Probabilistic record matching is a technique often used in conducting research studies. The probabilistic method 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. Probabilistic record linkage assumes that no exact match between fields common to the source databases will link a person with complete confidence. However, many researchers, including those at the Chapin Hall Center for Children, have used this method with great success.

Chapin Hall's core matching fields include: first and last name, birth date, SSN, gender, race, ethnicity, address and county of residence. Once a match has been determined, Chapin Hall researchers assign a unique number to the matched records. The completed record matching process creates what Chapin Hall calls a "master link-file," which contains the unique ID number, client identifying data, and all ID numbers assigned by agencies from which the client receives services. As a result, when a client receives services from multiple agencies, the various records belonging to the client are linked.

## Tips and Special Considerations

- **Confidentiality.** Creating unique identifiers using personal identification information may limit the use of that data due to privacy laws and restrictions on data sharing.

### *Example 1: Connecticut Public School Information System (PSIS)*

Privacy protection laws such as The Family Educational Rights and Privacy Act (FERPA) and the Privacy Act of 1974 restrict access to personally identifiable information in PSIS and place limits on how data may be used. Until the assignment of unique identifiers, which enables personally identifying information to be "de-linked" from a student record, PSIS data can only be used in aggregate form for compiling summative reports. Currently, record identifiers are composed of student names and birth dates.

### *Example 2: Massachusetts Office of Child Care Services (OCCS)*

The Massachusetts Office of Child Care Services needed to match their database on certified child care teachers with the wage and employment information from Unemployment Insurance. The Department of Revenue (DOR) is very cautious about releasing such sensitive data. OCCS is negotiating with the DOR to set up an ongoing data match. The match was done using a longitudinal database maintained at DOR and made available to outside researchers (Chapin Hall) in aggregate.

Contact: Rod Southwick, Office of Child Care Services, [rod.southwick@state.ma.us](mailto:rod.southwick@state.ma.us)

## ENHANCEMENT STRATEGY 2

### Establish common variable definitions and coding.

#### Overview

Variable definitions and coding can vary across and within agencies, making it difficult to link databases and answer a broader array of questions about programs and services. For example, one database might use “M” and “F” for male and female, while another may use numeric values – “1” for male and “2” for female. Incongruities such as these necessitate recoding to a common format before further analysis can take place.

In addition to impeding data linkages, variations in coding schemes can also limit the level of analysis that can take place. For example, one database might record an individual’s age using numeric range categories (e.g. 10-15, 16-20, etc.) instead of actual numeric values (e.g. 10, 11, 12). Another example might be recording “yes” or “no” in a data field capturing whether a child has participated in a preschool program instead of recording the actual type of program (e.g. Head Start, School Readiness, etc.). Collapsing or simplifying data in this manner compromises more in-depth analysis. Efforts to code and record data in a consistent and research-friendly format can eliminate hours of recoding and improve overall research capacity and database utility. Furthermore, standardized definitions make it easier to share data within and outside of the agency because there is no confusion about what different variables mean or measure.

#### Guidelines for Successful Implementation

1. *Survey existing data elements that are collected, their coding schemes, and definitions.* Document differences within and/or across agencies.
2. *Inventory information needs and determine what data must be captured.* The following information needs should be taken into consideration:
  - a. Program planning, evaluation and service delivery (both short and long-term)
  - b. Various reporting requirements for federal and state funding sources, legislative bodies and the general public
  - c. Research and policy development
3. *Catalog how data must be reported to different audiences* (e.g. annual or monthly data averages, specific age groupings, geographic regions, etc.).
  - a. Determine an appropriate unit of analysis for all variables that will satisfy both reporting requirements and support research and planning.
  - b. Determine any data programming needed to collapse more detailed data into specific groupings that are required for standard reports (e.g. consolidating 20 race categories into the 5 categories required on a monthly report submitted to the federal government).

In addition to impeding data linkages, variations in coding schemes can also limit the level of analysis that can take place.

### Tips and Special Considerations

- **Data collection flexibility.** Although federal agencies may specify how data elements should be formatted in reports submitted from the states, this should not govern how source data is collected and recorded. Technical applications can convert data extracts into the appropriate format for transmittal to federal agencies. However, as coding schema are developed, it is important to ensure that conversion is possible so that it does not hinder reporting requirements.

***Example:** Reporting on Child Care Development Fund Grants*

The Child Care Bureau in the federal Department of Health and Human Services requires states to report on families and children served with grants from the Child Care and Development Fund. The electronic file that is sent to the Bureau is in a specified format. States are required to draw the families and children served from their own databases and convert fields that are formatted differently into the Bureau's requirements. Each state produced its own conversion application so that this process can be done electronically before the file is sent.

Contact: Joseph Gagnier, Child Care Bureau, [jgagnier@acf.hhs.gov](mailto:jgagnier@acf.hhs.gov)

- **Caution: “check-off” variables.** A check-off variable refers to a data collection method whereby when the response box is checked, it means “yes”. This variable formatting can cause confusion, whereas an unchecked box might indicate a negative response *or* a missing value.

## ENHANCEMENT STRATEGY 3

### Implement and improve archiving functions.

#### Overview

Archiving past data helps one to assess changes over time, which is a critical component of program and policy evaluation. With an archive of past data, internal and external researchers can match cohorts of data across databases and conduct comparative and longitudinal analyses. Unfortunately, many administrative databases are not equipped with this function and new information simply overwrites existing data and eliminates a great deal of research potential. To enable these comparative and longitudinal analyses, all databases should contain archive functions and data should be saved systematically. When resources are not available to upgrade databases to include an automatic archiving function, another option is to implement standardized procedures for saving electronic files of past data and making them available for future analysis.

#### **Reaping the Benefits of Archived Data**

The Ray Marshall Policy Research Center at the University of Texas provides a database archiving service for state agencies that is not only helpful to the agencies, but also facilitates University research on public policy. Data sharing agreements still must be executed when the University wants to use the archived datasets. One study that was a direct result of having archived datasets was an analysis of the effect on child care subsidy program outcomes associated with policy changes coming from the devolution of federal programs to state administration.

#### Guidelines for Successful Implementation

1. *Convene a workgroup to formulate a strategic plan for archiving.* The following key issues should be addressed:
  - Will the archiving function be initiated automatically (i.e. computer generated) or will it require some level of staff involvement/manipulation?
  - How frequently will data be archived (daily, weekly, monthly)?
  - Will the entire database be archived or just certain data sets extracted from the database?
  - Once archived, how will one be able to access past data? Will access be restricted to certain users?
2. *If data is to be manually archived, implement adequate staff training to ensure that data is archived correctly and at regular intervals.*

## Tips and Special Considerations

- **Involve the research community.** In addition to agency staff, consider inviting external researchers to join the strategic planning discussion. Getting researchers' perspectives on how frequently to archive data and what data to archive can help shape a strong archiving process that not only meets the agency's operational needs, but also supports future research.

## ENHANCEMENT STRATEGY 4

### Reduce missing or erroneous data.

#### Overview

When critical data is not collected or when it is entered incorrectly, it greatly limits the potential of administrative databases to address key policy and programming issues. **Absence of data not only compromises which questions a dataset can answer, but it can also complicate record matching.** For instance, if “date of birth” is one of the fields used to uniquely identify a client during the record-matching process and a data entrant skipped over this field, the task becomes more challenging. Another problem arises when researchers attempt to conduct telephone surveys using administrative data and encounter outdated client contact information (see below for tips on staying current). In most agencies, not surprisingly, the amount and accuracy of the data in administrative databases is directly related to its usefulness in day-to-day operations. For example, if certain data is irrelevant to the immediate task, it is common for busy workers to skip over a data field entirely even if inclusion of that data might help answer some important questions that arise later on.

Updating data systems so that users cannot progress to a new field until data has been entered in the preceding fields is one way to address this issue. However, safeguards must also be put in place to avoid the entry of “dummy” values (e.g. \$1.00 for a full-time worker’s annual salary) to circumvent these quality assurance measures. For example, software that flags questionable entries such as the example listed above can help address this issue.

However, a better, or at least complementary, approach involves focusing on training and fostering a commitment to data quality among staff. It is important that data entrants not only sharpen their technical skills, but also understand the underlying value of quality information to both agency operations and policy making. Source data is more likely to be reliable if the person entering the data understands the ramifications – broadly and individually - of poor data quality. Workshops and consultation sessions that illustrate these points can help establish a workplace that values *and* produces quality data.

#### Guidelines for Successful Implementation

1. *Actively engage caseworkers/data entry staff during database or system development.* Assessing their information needs and work flow processes and incorporating these into the design increases the probability that staff will be willing and able to comply with the agency’s data entry improvement efforts.
2. *Train and educate agency staff on new database functions and data entry procedures, as appropriate.* Be sure to underscore the benefits of increased data reliability for the staff member, agency and others. Work to expand the agency mindset from a sole focus on day-to-day operations toward a vision for impacting the future.

3. *Conduct a periodic evaluation of the database structure (i.e. ease of use, compatibility with work flow) and data quality to identify areas for improvement. Solicit feedback from those entering and using case data as well as external researchers who worked with the agency's administrative data.*

## Tips and Special Considerations

- **It pays off.** Although the initial costs of training may be high, the return on investment is substantial over time.

### **Tips on Keeping Client Records Current**

*(Excerpted from Mathematica Policy Research, Issue Brief Number 1, July 2002)*

- Update automated records frequently.
- Include the complete address – street address, apartment number, post office box, and accurate zip code.
- Include the telephone number. Ask people without telephones for “message numbers.”
- Give clients postcards to mail in address or telephone number changes.
- Merge/coordinate with other agencies that update addresses more often.
- Use “Address Service Requested” for mailings.\*
- Keep area codes up-to-date and revise when new area codes are added.
- Encourage clients to provide unlisted telephone numbers.

*\*Note: With this notation, the post office attempts to forward mail for a specified time period and the sender is either notified of address correction or reason for non-delivery. Check with the US Postal Service for associated fees and restrictions with this service.*



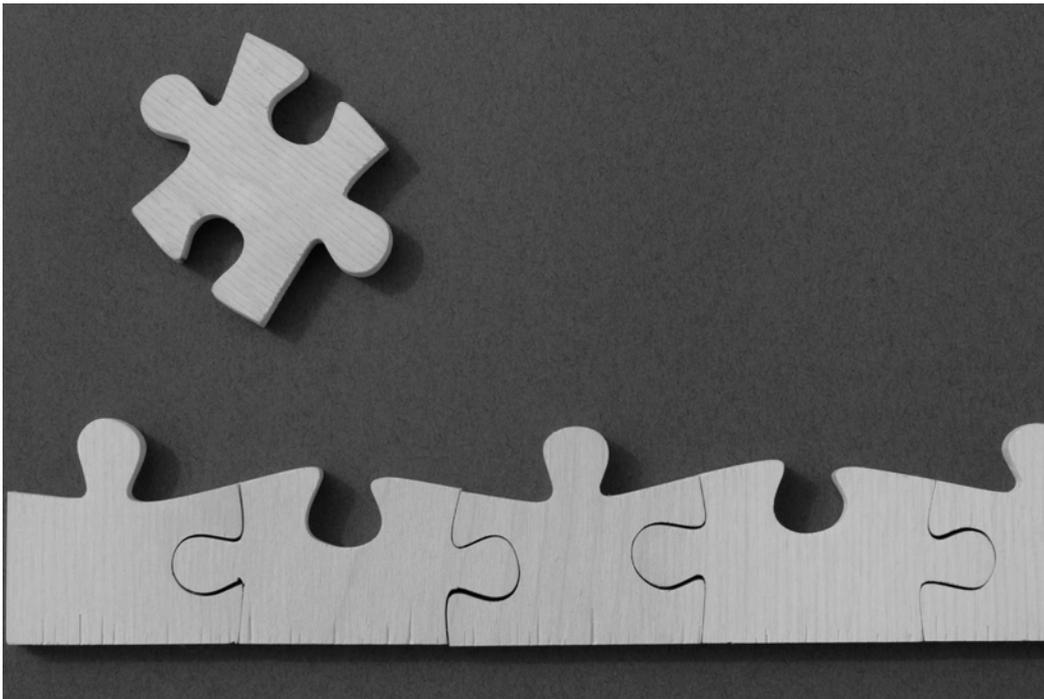
## Making Connections: *Data Linking/Integration Enhancements*

---

In many instances, generating the answers to key policy and planning questions using administrative data requires linking or integrating multiple databases. Making these connections exponentially expands the utility and informative power of administrative data, supporting the exploration of more sophisticated questions. Laying a foundation of quality data, which includes components such as common record identifiers and common variable coding, can facilitate these data connections.

This section outlines the two primary strategies for connecting data across various sources:

- Linking individual databases or data sets (in an ad hoc manner)
- Integrating multiple data sources (in a more complex manner)



## ENHANCEMENT STRATEGY 5

### Link individual databases or data sets.

#### Overview

Data linkage is the process of combining two or more unique databases/sets in an ad hoc manner into a meta-database by matching unique record identifiers within the databases. When data is linked in this manner, or through more complex data integration procedures (see Enhancement Strategy 6), it provides a much richer data set with greater capacity for research and analysis. Data linkages can help state agencies learn about:

- Overlap in client population
- The extent to which transition or referral takes place from one program to the other
- The correlation of outcomes with conditions, program participation or interventions
- A broader range of information on characteristics or conditions of a target population

Linking data exponentially expands the utility of administrative data, which is why much of this toolkit is focused on getting data in shape for data linkages and actually pursuing data linkages or data integration. Data linkages may involve data from across programs and/or agencies, as well as linkages between agency data and survey data. Each linkage adds another layer of analysis and information to support decision making. Several basic elements covered in this publication that facilitate data linkages include establishing common record identifiers and comparably defining and coding variables within and across agencies.

#### A Sampling of Data Linkages

##### **Connecticut Department of Public Health and Children's Health Council**

Data Linkage: *Birth registry data and Medicaid eligibility data*

Outcomes: Integration of these two datasets enabled researchers to examine the relationship between prenatal conditions and birth outcomes for women whose care is funded by managed care programs.

##### **Connecticut Departments of Education (SDE) and Mental Retardation (DMR)**

Data Linkage: *Integrated Special Student Information System (ISSIS) and Birth-to-Three Early Intervention Program*

Outcomes: This annual data linkage provides the means for longitudinally studying the long-term impact of early identification and service interventions on children with special needs as they transition from Birth-to-Three to Preschool Special Education and into elementary school.

##### **Oregon Employment Department, Child Care Division (CCD)**

Data Linkage: *Child care licensing, subsidy, and resource and referral data, and Child and Adult Care Food Program data.*

Outcomes: Through an annual merge of these databases, Oregon is able to determine what proportion of providers are actually participating in the subsidy program. Currently, the merge is using a probability matching methodology. Partners are moving to a common identifier to expedite merges. For more information, contact Roberta Weber, [bobbie.weber@oregonstate.edu](mailto:bobbie.weber@oregonstate.edu).

## Research Benefits of Potential Data Linkages

Linkage: WIC records and Pregnancy Risk Assessment Monitoring System (PRAMS) survey data

Research Benefit:

- Comparative analysis of breast-feeding practices could be made for both WIC and non-WIC participants.

Linkage: WIC and state Medicaid eligibility files

Research Benefits:

- An increased ability to assess the extent of unmet need for WIC and Medicaid services on a local level;
- The capability to estimate the degree to which cross-referrals can be made among WIC, Medicaid and prenatal care programs and their potential impact on infants' health;
- The ability to identify mothers who do not enroll their children in the state's immunization registry; and
- The means of comparing Medicaid and non-Medicaid populations on adequacy of prenatal care and birth outcomes.

Linkage: Public School Information System (PSIS) and various early childhood databases

Research Benefit:

By linking a child-level early childhood database to PSIS, which contains child-level data for students grades K-12, one can examine how children with varying early childhood interventions.

## Guidelines for Successful Implementation

1. *Evaluate the quality of the datasets to be linked using the following criteria:*
  - Consistent values can be established for the data fields used for matching
  - Variables can be broken into components that need to be compared
  - Typographical errors can be identified and rectified
  - De-duplication procedures can be employed
2. *Create a record-matching plan that includes the following components:*
  - Description of the data files to be linked
  - Selection and description of the data elements to be matched – making variable attributes consistent across files – identifying cleanest and most unique variables
  - Procedures for standardizing data elements across systems and over time – recoding categorical data elements and standardizing the structure and content of data fields so searching and comparing can be effective
  - Procedures for ensuring data retention – retention of information when individual data fields are updated to reflect changes
  - Record layout – description of where data are located in the data file
  - Format of the new matched data file
  - Process for creating a Unique ID for matched records

- Algorithm for matching records – determination of algorithm structure to be used:
  - Single-field comparison – individual field chosen for match and all records clustered together with similar values for that field
  - Multi-field matching – multiple fields from two records are compared and an aggregate matching score computed by combing the results of the individual field comparisons
  - Rule-based matching – multi-field combinations are executed and then applies a set of if-then, condition-action statements to match records
- Specification of steps for matching records - delineate combination of variables to be matched in each successive step – apply cleanest and most unique variable first
- Process for verifying matched records and specification of software for addressing anomalies in identifying information, i.e., misspellings, missing letters, surname – automatic or manual methods for correcting false-positive and false-negative errors – completed after each matching step -
- Procedures for cleaning data – how mistakes in matching records, i.e. duplicates, will be undone

## Tips and Special Considerations

- **Cover all bases.** When linking data across agencies, be sure to seek the counsel of both agencies’ legal/contract unit or other designated entity to ensure that confidentiality and legal restrictions around data sharing are adequately addressed. Representatives from both agencies need to work together to develop mutually accepted Memorandum of Understanding (MOU) language.
- **Don’t forget survey data.** Linking administrative and survey data (e.g. US Census) enables more in-depth analysis, whereas each data source contributes a different level of detail. For example, “...while administrative data provide detailed information on the characteristics of program participants, the services they have received, and the actions they have taken, they are not useful for estimating such things as the rates of program participation because they do not account for the entire population eligible or at risk for the program.”<sup>5</sup>

“Linking information from administrative and survey data sources appears to offer the best of both worlds...”

~ Joint Center for Poverty Research<sup>6</sup>

## ENHANCEMENT STRATEGY 6

### Integrate data.

#### Overview

The most ambitious and multi-functional technical “upgrade” is the integration of multiple data sources within or across state agencies into a single, reliable platform. In many cases, this platform is a data warehouse. A data warehouse is a meta-database resulting from a data integration process by which data from multiple sources is cleaned, put in a standardized format, and merged so that it can be fed into the warehouse. Although data warehouses are typically “read only”, they provide a user interface that enables users to query the system and generate reports and data marts, which contain subsets of information extracted from the warehouse.

Data mart: a subset of information extracted from a data warehouse.

A data warehouse (and the integrated data within it) is considerably beneficial because it:

- Provides timely and accurate responses to user queries
- Enables users to query across database boundaries
- Provides access to data that has been converted from varying sources into compatible formats for linking and analysis
- Supports longitudinal analysis and analyses across programs and (potentially) service settings/agencies
  - Helps portray individual experiences of clients as they enter and exit the service system
  - Helps determine effectiveness of services when there are multiple interventions and needs for coordination of services
  - Helps measure the impact of policy changes

Effectively integrating data and, potentially, developing a data warehouse is a challenging task that requires thoughtful planning and an ongoing commitment from a variety of stakeholders, including executive leadership, IT staff and end users. Many concerns are likely to surface during the process, especially if the data being integrated stems from multiple agencies. Some of these fears include, losing control of programmatic databases, sacrificing division personnel time to the project, negative impacts on service delivery. Data integration proponents may also encounter concerns over legal (e.g. confidentiality) and political issues or cynicism that the project will ever be successfully implemented. Leadership backing is essential to help overcome these fears.

Hopefully the guidance outlined in this section will help minimize setbacks. The guidance is based on “lessons learned” from various states that have pursued data integration and data warehouse development efforts. Guidelines for this section are divided into two parts: the first is broadly applicable while the second is more specific to *interagency* data integration/data warehouse efforts. The guidelines previously discussed in this toolkit can also help, whereas many of the enhancement strategies addressed represent foundational elements that support data integration, particularly those that assist with linking data and improving data quality.

## **Top Ten Reasons Why... When You Hear Data Integration, Come to the Table**

*(excerpted from a presentation by Garland Land, Missouri Department of Health and Senior Services)*

1. IT resources are maximized through consolidation
2. The up front investment seems high, but overall maintenance is lower
3. An integrated system reduces data collection burden
4. Expansion of an integrated system is easier than developing a new categorical system
5. End user training is simpler
6. Data integration can improve service delivery
7. Data integration can improve efficiency
8. Data integration can tear down categorical thinking
9. Data integration can drive better program planning and policy development
10. Data integration provides client centered outcomes versus programmatic outcomes

### **Guidelines for Successful Implementation**

*Note: Although these guidelines explicitly discuss development of a data warehouse, much of the guidance is generally applicable to data integration efforts.*

#### **Phase One: Project Planning**

1. *Secure financing of start-up costs.*

Although funding is important to any project, it is especially critical to establish financing early on for a project of this magnitude. Start-up funding ensures that the project will be able to purchase the full-time technical expertise required to successfully develop a data warehouse. This expertise is essential because the responsibilities that come with project development and implementation are substantial and far outstrip the capabilities of staff who are only able to devote a portion of their time to the project. Evidence in other states has shown that these types of initiatives are likely to fail when they lack sufficient staff resources and experience to keep the project moving forward.

Early financial commitment also brings an intangible, but essential, element to the project - political commitment. By putting their funds behind the project, agency executives signal its importance throughout the organization. The experience in other states has also shown that the willingness of agencies to commit funding at the outset usually results in ongoing support as the project expands and evolves.

2. *Develop a more complex steering committee structure.*

As previously discussed, steering committees are important to keeping any project on track. When undertaking a project of this scale, it may be necessary to expand the committee to include sub-committees that are responsible for tackling specific issues or task areas related to project development. For example, one should consider establishing a sub-committee to manage multiple funding streams, whether across agencies or across

programs within one agency. It is also critical that the full committee or a sub-committee analyze issues related to confidentiality, privacy and gaining access to data.

3. *Conduct an information assessment study.*

It is critical to identify all of the data sources to be included in the data warehouse, as well as data flow processes and data issues before any business applications solutions are considered and any data is integrated into the data warehouse. The successful operation of a data warehouse is dependent on identifying where variables need to be standardized, duplicated data eliminated and unreliable data suppressed, and how these functions will be achieved.

The first step in gaining a thorough understanding of where databases/data sets are comparable and divergent should be to identify and compile descriptive information on each data source within the organization. This information could include:

- Data type
- Acronym/ State Reference
- Data collection methods
- Time period covered
- Unit of analysis
- Description of population included
- Total number of records/ cases
- Identifiers for matching
- File locations
- Data limitations

The next step involves the identification and visual depiction of the business processes, activities and data movements between the data sources. Basically this requires a business unit-by-business unit analysis of all of the data sources utilized and how they flow into, out of and within the unit. Structured Analysis and Design Methodology, developed by Ed Yourdon and Tom DeMarco<sup>7</sup>, is one methodology that has been used to complete this function. The methodology includes the following elements:

- **Process.** These are business activities or functions that receive, process or manipulate information.
- **External Entity.** This represents a person or organization that is a supplier or receiver of information that is supplied into processes.
- **Data Store.** This is a place where data are stored.
- **Data Flow.** This represents the movement and direction of data from one place to the next.

In addition to inventorying data sources and illustrating how they flow through the organization, a very important aspect of the study is to ascertain the organization's information needs and the factors that negatively impact them. These insights are gained through in-depth interviews with key staff members responsible for managing the data sources, senior agency leadership, field workers entering data on-line and contracted service providers.

This information gathering process should focus on the following issue areas:

- **Data.** This area encompasses a broad scope of data utilization processes including:
  - Information needs for service delivery and program planning and evaluation
  - Reporting requirements for federal and state funding sources, legislative bodies and public inquiry
  - Data entry and storage, and overall data quality
  - Queriability of data sets and capacity for linking data sets
- **Process/ Operational.** The focus here is on identifying operational processes that inhibit the optimal use of data, including:
  - Policies that result in duplicated data collection
  - Independent databases dispersed throughout the agency
  - The influence of service delivery contractors on how data is collected and managed
  - A lack of coordination between business units for data utilization
- **Communications.** The extent to which communication (or lack thereof) adversely impacts the optimal use of data should also be studied. Among the areas of interest are:
  - Procedures for submitting requests to IT managers for data extracts
  - Capacity for sharing data across program units and agencies
  - Capacity to respond to data requests from outside groups, including university researchers, legislative bodies and policy organizations
  - Efficacy of communication channels between senior management and program staff in supporting data management and maintaining data quality

The final step of the assessment study is to formulate recommendations for improving data management and utilization within the organization through the data warehouse.

## **Phase Two: Data Warehouse Design**

1. *Utilize the experience and insights of data warehouse design experts.*

Many organizations lack the internal expertise to develop an optimal warehouse design to meet their data needs. This is not surprising given the broad array of system components – from matching individual client records to creating a user-friendly interface. Consideration must be given to different methodologies, database software programs, operating system platforms and, ETL (Extract, Transform and Load) and reporting tools. Some organizations have therefore turned to outside experts to assist them in designing their data warehouse.

Experienced data warehouse designers help organizations to translate their informational needs into functional requirements. They take the lead in developing application specifications, selecting software tools and evaluating system performance. Most importantly, they have the expertise to present a variety of options and to assist the steering committee in making the proper selection.

2. *Design the data warehouse and select technology to meet desired business processes and informational needs (i.e. match technology to identified user needs).*

Generally there are four primary components that make up a data warehouse:

- data acquisition
- data management
- data delivery
- system management.

For each component, processes must be designed and Business Intelligence (BI) tools selected to support their function. As a rule of thumb, the decisions surrounding these processes and tools should be guided by the organization's informational needs as identified during the assessment study. Business information goals drive function requirements. Required functionality, in turn, directs decision making on business processes and the tools that support them. Experienced data warehouse designers can help agencies determine which BI tools can best meet their needs.

***Example:*** *Massachusetts Office of Child Care Services (OCCS)*

The Massachusetts Office of Child Care Services built an early care and education data warehouse to integrate data sets from their office and other state agencies, as well as federal census data and state surveys. The project was designed to inform child care policy issues and allow child care stakeholders to access numerous sources of data and address longstanding child care issues. Contact: Rod Southwick, Office of Child Care Services, [rod.southwick@state.ma.us](mailto:rod.southwick@state.ma.us)

3. *Establish procedures that ensure data quality and data warehouse integrity.*

Data quality issues are a major concern in the operation of an enterprise-wide data warehouse. For the most part, data warehouses consist of multiple systems that contribute data in different formats and standards and with different contextual meanings. As a result, data quality is frequently corrupted by mistakes, missing data and lack of standards with respect to variable definitions and coding. Considerable time must therefore be spent on developing procedures and selecting tools that ensure data accuracy, completeness and relevancy.

Several of the enhancement strategies mentioned previously in this toolkit address data quality issues (e.g. establishing common variable definitions and coding). Following through on these quality-enhancing strategies with each contributing database prior to embarking on development of a data warehouse can facilitate implementation. However, there are other methods for dealing with data quality in data warehouses, such as the establishment of data marts and/or utilization of Data Quality and Integration (DQI) tools. Data Marts are derived from a data warehouse, but include a more defined set of data that is relevant to the specific needs of a department or a particular set of users. A data mart's narrower focus allows for more precise monitoring and management of data quality. Some states however have determined that the use of data marts alone is insufficient for maintaining high quality data. In these cases, DQI tools have been incorporated into the system design.

DQI tools integrate, standardize and de-duplicate data. They are designed to work in conjunction with developed enterprise standards to ensure that the data entered are valid and complete. The front-end data validation check is only one aspect of the tools. In a data warehouse environment, it is equally important to deal with the migration of data to a new database or platform. A DQI tool therefore also works to accurately identify the source data, transform it into the proper format and move it to the desired database.

In lieu of (or in addition to) using these methods, some states have developed guidelines to govern maintenance of data quality. For example, to limit duplicative data and ensure data quality in their FamilyNet integrated health data system, the Oregon Department of Human Services instituted the following business goals:

- *Maximize data quality.* No more than 5-7% duplicates are allowed in the Client Master database. Manual cleaning of the data may be required even after automated best practices are implemented. Copies of useful old demographic data (address, phone numbers) are maintained.
- *Define clear criteria for identifying duplicate data at the Client Master and Module Level.* Client Master data are first priority. Module data are second priority.
- *Create an import utility to be used for all Client Master Data Imports.* Each FamilyNet module will be responsible for the process to import, de-duplicate and merge module data.
- *Create a batch process to review Client Master data.* The process should identify duplicates and merge duplicate records based on specific match/ merge rules.
- *Create an efficient on-line process for reviewing and resolving instances of duplicate Client Master data.*

### **Phase Three: Data Warehouse Implementation**

1. *Use a phased-in implementation schedule.*

Several states building enterprise-wide data warehouses have had success in taking an incremental approach toward implementation. This phased-in approach can help planners to overcome many of the technical and financing issues that may arise as the project rolls out. Further, it allows interim successes to serve as the base upon which support can be built for subsequent expansions.

Planners in many states have deliberately scaled back the scope of system applications early on. They have pilot tested their system designs and data integration and reporting tools using selected data sources. This has helped them to spread out development costs over time, make refinements, build infrastructure support and gain expertise in the data analysis and reporting tools.

2. *Develop metadata and other tools to assist users in maximizing the system's potential.* The experience of many states is that the users' ability to employ the data warehouse to its fullest capacity is dependent on their level of understanding of how the system operates and their knowledge of its constituent parts. Lack of easy access to metadata, less than adequate training in using it and lack of tech-

**Metadata:** Generally defined as data about data. Metadata can include data element descriptions, data type descriptions, attribute/property descriptions and process descriptions.

nical support can lead to under-utilization of important functions. In addition to basic technical support, such as a printed training manual and the built-in software Help function, other essential tools include off-site and on-site software training, telephone/ help desk support and on-site technical assistance.

***Example:*** WISDOM (Wisconsin)

In Wisconsin's WISDOM system, a metadata application was designed to increase users' familiarity with the data. Information involving every data element in WISDOM, both business and technical metadata, was made available to end-users via the Metadata application. In addition, users who want to include a given data element in a query can use the Metadata application to learn the business meaning of that element and track it and all the transformations that created it through the Oracle database element, the mainframe extract and the source system.

#### **Phase Four: Data Warehouse Administration**

1. *Identify a full-time data warehouse administrator.*

A full-time position is necessary to oversee the management and monitoring of the data warehouse. A variety of problems are certain to arise as the system gets up and running. It is critical that one person serve as the acknowledged clearinghouse for addressing emerging issues and fielding complaints from users. Importantly, the administrator should be given the authority to ensure that adequate resources, physical and financial, are quickly deployed to resolve problems. The administrator should also work cooperatively with program managers and IT staff to continually review the interoperability of the system's hardware, software and data management technologies.

2. *Develop measures that track return on investment.*

It is imperative for organizations to measure the impact that the availability of quality, integrated data is having on its business operations. Accordingly, the organization's stated business goals need to be transformed into quantifiable terms that can be measured and evaluated. Tracking return on investment is an effective way to justify and ensure continued growth of the warehouse.

3. *Develop policies that specify data warehouse administration and data sharing guidelines.*

Special guidelines should be established for when and under what conditions data will be made available to users. Such guidelines should be perceived not only as a means for governing the use of data enterprise-wide, but also for encouraging optimal usage of the integrated data made available by the warehouse. In addition, guidelines should be developed that define the requirements for long-term maintenance and archiving of the data (see item 5 below).

4. *Assume the philosophy that the system will continue to evolve and to require on-going maintenance.*

The experience of other states has shown that change is a constant in the operation of a data warehouse. The interoperation of the system's design components will not always function as desired. Additional data sources will be added and new functions will be integrated based on emerging informational needs. As a result, new applications will have

to be developed and new tools selected. The continual evolution of the data warehouse environment necessitates that the deployment of staff time and expertise, outside consultation and financial resources be ongoing.

5. *Establish a system maintenance plan.*

A plan should be in place that specifies ongoing procedures for maintaining the system's functionality and performance. Standardized processes for reporting problems should be set up and communicated to end-users. In addition, a schedule for diagnostic review and testing should be established. Normal maintenance requirements generally include rectifying system "bugs", updating reference tables, installing upgrades to architectural components and making system modifications to meet mandates contained in new federal and state governmental policies, procedures and reporting requirements.

It is also critical that processes are put in place to help keep client and agency data current. The system is valuable to users only if the client and agency data is accurate, complete and up-to-date. Caseworkers encountering outdated or inaccurate data quickly become frustrated, leading to a potential breakdown in the data entry process.

## Additional Implementation Guidelines for Interagency Efforts

The following guidelines are repetitive of other sections, but are worth highlighting because of the size and complexity of the undertaking and because databases may be included from several state agencies.

1. *Convince the executive leadership of the various state agencies of the benefits of participating in an interagency data warehouse.*

Buy in from commissioners, their executive staff and managers to integrate their databases with other state agencies is fundamental in overcoming the numerous obstacles that emerge when an interagency data warehouse is contemplated. The reasons to opt out are many and varied. There are concerns over client confidentiality and privacy, and the potential legal actions that could surface. For some agency executives, the sharing of data equates to a loss of control over the data. This perception gives rise to an uneasiness that the data may be used in ways that come back to embarrass them. Furthermore, agency executives correctly realize that such an effort comes with a substantial price tag for financial and staff resources.

At the same time, however, there is likely a realization that good data is critical to fulfilling the agency's mission. The goal for data warehouse proponents, then, is to address the political and legal concerns up-front (see below for additional guidance) and to clearly articulate the goals, uses and benefits of an interagency data warehouse. One convincing argument is that a data warehouse provides an opportunity to obtain accurate, timely and comprehensive information on client populations and services that can in turn help administrators improve policies and service delivery. Giving concrete examples of how integrated data have been used to benefit other agencies may also help to bring administrators on board. The documented experience of other states can be effective for portraying the positive outcomes realized through the use of integrated data as well as for alleviating fears that the privacy of clients will be exploited.

For commissioners, the business equation will be to assess whether the returns from a data warehouse will be greater than the costs and possible risks incurred. The first and most daunting task for proponents will be to convince them that it is indeed a cost-effective investment.

2. *Address confidentiality and privacy issues as well as any restrictive legislation or political issues regarding data sharing early in the project's development.*

As mentioned above, some agency leaders will have serious concerns that integrating data sets increases the risks of disclosure of personal information. The central reason for this concern is that the combination of databases will provide more identifying information than any of the separate files. Furthermore, there is apprehension that increasing user access to identifying information invites problems involving personal privacy. It is essential that agencies feel comfortable with the security and confidentiality policies and protocols so they are willing to release their data without expunging the identifying elements that are critical to record matching. To help agencies reach this level of comfort, data warehouse proponents must demonstrate that the system can maintain high standards of confidentiality and securely guard access to its data. In designing an inter-agency data warehouse, confidentiality and privacy issues within each agency must be considered seriously and the safeguards required by protective legislation must be addressed.

Some steps proponents might propose include:

- Using informed consent forms to allow clients to determine who has access to their information, what information is accessible and for what purposes the data may be used
- Assigning passwords to various users or implementing other security measures to control levels of access
- Monitoring patterns of data access through audit trails

In addition, it is helpful to present steps that similar projects in other states have taken to minimize the risks of confidentiality problems.

***Example:*** *KIDSNET – Rhode Island Department of Health*

Although Rhode Island's KIDSNET data warehouse is not an interagency effort (it links data from several different programs within one agency), it provides valuable insight on addressing confidentiality issues. Rhode Island's approach to dealing with confidentiality was a long, deliberative process resulting in an approved policy governing access to information. Thoughtful research and consensus building were important to the policy's development, including the following key steps:

- a. Review of program specific laws
- b. Integration of "Need to Know" policy
- c. Involvement of program managers and other stakeholders to identify "who should have access to what data"
- d. Input of parents
- e. Development of a Provider Agreement; Policy Handbook; and Notification Process

- f. Development of a process to request approval by programs for any new requests to access their data through KIDSNET.

It is essential that agencies feel comfortable with the security and confidentiality policies and protocols so they are willing to release their data without expunging the identifying elements that are critical to record matching.

Existing state agency policies that restrict user access and govern the use of administrative data (like Rhode Island's, above) should be collected and analyzed. Similar policies from states that have established interagency data warehouses should also be obtained. A summary of their salient aspects should be incorporated into the initial presentation made to senior agency leaders. This toolkit can help proponents begin this investigative work.

A similar approach should be undertaken with regard to institutional review boards (IRB). IRB policies currently employed by state agencies should be examined and summarized. Their ability to protect individual privacy and eliminate the need for repetitively obtaining informed consent should be spelled out. Also a prototype data sharing agreement that outlines usage restrictions and clearly exhibits how the research will be applicable to state agencies (e.g. research will include a focus on service improvements) should be part of the initial presentation.

Data warehouse designers will also have to have a firm grasp of how regulatory legislation such as HIPAA (Health Insurance Portability and Accountability Act) and FERPA (Family Educational Rights and Privacy Act) impacts access to and use of integrated data.

## Tips and Special Considerations

- **Be prepared for challenges in securing project funding.** Some states have indicated difficulties in obtaining financial commitments from programs, departments and/or agencies because they objected to the fact that other entities that were not contributing equally or at all would also benefit from the data warehouse.
- **Source data is key.** When designing the data warehouse, remember that source data are most reliable when they are integral to the system's purpose, i.e. data quality is better if it's useful for program functioning and not just research. Consider both of these usages throughout project development.
- **Going online: a word of caution.** Some states have found that putting the system online can reduce its capabilities because web deployment can limit available functionality to only those functions that are supported by widely used web browsers.

- **Optimize data analysis.** To facilitate more in-depth analysis using linked or integrated data, try pursuing some of these techniques which are used by researchers at Chapin Hall:
  - Create summary records that represent significant periods in clients' service histories. These summary records can assist in studying patterns of duration and re-entry into service systems as well as analyzing the timing of service interventions and long-term patterns of service usage.
  - Conduct geographic analysis using information such as mailing address and county of residence. This "geocoded" information can be used to construct profiles of service provision and client composition in a given city or region.



## Sharing with Others: *Data Access Enhancements*

---

“We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health.”

~ National Institutes of Health (NIH)  
Final Statement on Sharing Research Data<sup>8</sup>

Improving external researchers’ access to state administrative data is a mutually beneficial arrangement for both parties. In partnering with universities and policy organizations, state agencies obtain performance assessment data and outcomes-based research that can help increase the effectiveness of their practices and lead to more responsive policies. For researchers, administrative data is essential in examining critical human service issues using scientific methodologies. Using administrative data, researchers can match client records and create integrated data sets thereby enabling them to track individual clients over time, assess intervention outcomes and perform trend analysis.

However, there is some trepidation on the part of both researchers and state agencies in entering into these data sharing arrangements. State agencies must be concerned about how the data will be used and whether the uses are compatible with their goals, including the protection of client confidentiality. They have a well-founded apprehension about public scrutiny of their efforts and worry that researchers will not understand or be sympathetic to their resource or policy limitations. As with all research endeavors, state agencies have a keen interest in how the findings will be interpreted and whether they will have adequate opportunities for reviewing the findings before they are made public. On the other side, researchers need cooperation from state agencies in order to gain a clear understanding of complexity and quality of the administrative data they receive and to determine whether it will support good research. Researchers need assurances that the data are timely, accurate and complete, as well as access to someone in the agency who can answer questions as they arise during the data analysis and research process.

The first and most important step in addressing these concerns is establishing a solid working relationship between state agencies and outside researchers that is built on trust and mutual commitment. While much of this is dependent upon cultivating relationships, establishing tools and protocols that help address many of these issues (e.g. confidentiality, data transparency) can also help facilitate the process. Several of these tools are discussed in detail in this section, which includes the following enhancement strategies:

- Establish Institutional Review Boards (IRBs)
- Develop research protocols
- Develop user-friendly codebooks and data dictionaries
- Develop formal data sharing agreements

Below are some general guidance on relationship building among state agencies and researchers.

---

## Tips for Fostering Good Working Relationships Between State Agencies and Researchers

### **For State Agencies:**

- Develop a policy that articulates the agency's value for practice and policy relevant research, the importance of administrative data to such endeavors and the role of outside research organizations in assisting the agency in meeting its information needs.
- Identify the agency's research priorities on a three to five year planning cycle.
- Establish protocols that facilitate the process for conducting research with the agency, from proposal development to data acquisition and dissemination of the findings.
- Consider outside researchers as important stakeholders in the agency's efforts to achieve best practices and responsive policies, and include them as partners in policy and planning initiatives.
- Ensure that adequate staff resources are made available to facilitate the data collection of all agency approved research projects.

### **For Researchers:**

- Articulate how the agency's administrative data will be used and, to the extent possible, adapt the research design to accommodate the state agency's specific information needs.
- Provide the state agency with a proposal that fully meets all IRB requirements.
- Describe how the state agency will benefit from the data sharing arrangement and what new and relevant information they can expect to gain from the proposed research.
- Ensure that the state agency will have the opportunity to review and provide input on research findings before they are made available to the public.
- Adopt the philosophy that the agency's administrators and program staff have credibility in their interpretation of research findings.
- Meet with program analysts to develop a better understanding of their program operations, data collection procedures and reporting requirements.
- Participate in agency-sponsored research partnerships, forums and roundtables.
- Build a track record of research projects that improve practices and policy-making by providing solid information.

Another strategy for cultivating relationships is to "set a table" for discussion between state agencies and researchers. For example, in June 2003 Early Childhood DataCONNECTIONS convened an Early Childhood Research Roundtable at which state agency staff, researchers and other stakeholders met and explored together strategies for developing policy research partnerships. To download the proceedings, visit [www.chdi.org](http://www.chdi.org).

# ENHANCEMENT STRATEGY 7

## Establish Institutional Review Boards.

### Overview

Central to increased utilization of state databases is an Institutional Review Board (IRB) or its equivalent that will implement the agency policy. An IRB's primary purpose is to protect the rights of human subjects who are involved in research studies. It does so by ensuring that the proposed research design and procedures meet the appropriate federal regulations, state laws and agency policies concerning privacy and confidentiality. Some state agencies have historically had an IRB because of federal requirements, but it is a practice and structure that each state agency should adopt with clearly written guidelines to govern it.

In addition to regulating privacy and confidentiality practices, an IRB also provides other important benefits to the agency. First, the process a state agency undertakes to develop IRB guidelines helps build consensus on the best ways to govern administrative data availability and usage. Secondly, the dissemination of IRB guidelines provides researchers with a firm understanding of what is expected of them in requesting access to the agency's data. This guidance helps researchers to develop proposals that address the agency's requirements and to mitigate possible misunderstandings between the parties. Issuing guidelines also signals openness on behalf of the agency toward making their data more readily available for research purposes. Finally, the IRB review process can help researchers to re-structure their research in ways that are more in line with the agency's information needs. The end result is a greater number of well-conceived research projects that make important contributions to the agency's practices and policies.

### Guidelines for Successful Implementation

1. *Establish clearly written guidelines for the IRB that address:*
  - The control the agency will maintain over the use of their data
  - The steps that must be followed to obtain permission to use that data
  - What exactly can be done utilizing the data
  - What information researchers can disclose about the state agency operations
  - The protocols that will be used to protect the rights, privacy and welfare of human research subjects
    - This includes examining under what conditions researchers will be granted access to individual client data, which is required for in-depth statistical analysis.
2. *Establish a proposal format to guide researchers' submission of proposed studies that requests, at a minimum, the following information:*
  - Principal investigators
  - Funding source(s) and amount(s)
  - Project description
  - Objectives of the study
  - Hypotheses or research questions

- Scientific design
  - Methodology
  - Subject population
  - Interventions/ study procedures
  - Benefits and risks to the study subjects
  - Recruitment procedures
  - Consent procedures.
3. *Convene the IRB regularly to review submitted proposals.* The review process should include the following considerations:
- Is the participation of all subjects voluntary?
  - Are subjects' rights protected?
  - Is the consent process outlined in accordance with federal regulations?
  - Do the potential benefits to agency practices and policies outweigh the potential risks to the subjects?
4. *Periodically review the IRB procedures.*  
Evaluate whether they are effective in meeting the agency's goals in light of the proposals that have been reviewed and the agency's desire to promote useful policy and program research.

### Tips and Special Considerations

- **Ensure easy access.** IRB guidelines should be easily obtained from the state agency, including posting on the agency website.

## ENHANCEMENT STRATEGY 8

### Develop research protocols.

#### Overview

Unless they have established relationships with departmental staff, researchers typically find it very difficult to approach a state agency about a prospective study. They lack information about contact persons, the scope of administrative data maintained, the requirements for submitting a research proposal, and agency limitations on approved research. Upon project approval, researchers face another set of obstacles regarding access to the data and interpretation of the extracts they receive. These difficulties contribute to researchers' perceptions that state agencies are resistant to sharing their data and unwilling to study critical issues.

However, state agencies that establish a standard protocol for conducting research with agency data can minimize these perceptions. By standardizing their research process and making the information readily available, state agencies will engender trust and goodwill among the research community. In addition, they will ensure that the projects they are asked to consider will conform to agency specifications.

#### Guidelines for Successful Implementation

1. *Assemble a new or designate an existing committee to develop the research protocols.*  
The task of developing research protocols can be assigned to the agency's IRB, a research committee or an ad hoc working group. Representation should include agency staff with research, database and program expertise. Outside researchers can be helpful as well because the protocols are directed at them.
2. *Integrate the following components into the research protocol:*
  - Step-by-step procedures for submitting a research proposal, including a standardized format and submittal forms
  - Standardized forms for the monitoring and reporting of research activities and preliminary findings
  - Contact persons by bureau, program or service area to serve as entry points to the agency and to offer guidance and direction to researchers on how to best proceed with a proposed research project
  - Standardized format for memoranda of agreement (MOAs)
  - Policies that govern privacy and confidentiality, agency review of findings, and reporting of results
  - Policies that govern the disposition and ownership of data, including newly created data sets, upon project completion
  - Security requirements for the storage of data and controlling access to it
  - Client consent forms for obtaining permission to utilize case sensitive information
  - Agency approved procedures for physically or electronically accessing agency data
  - Abbreviated data dictionaries that contain user-friendly definitions of commonly used data elements

- Agency limitations on how their administrative data can be used

## Tips and Special Considerations

- **Give options.** 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. If a standard (i.e. interagency) protocol were developed, it should offer optional language for different agencies on certain topics. Also, once an agency protocol is developed, don't forget to incorporate it into signed agreements between the agency and various research organizations.
- **Dedicate a website or web page.** Try to disseminate information to the research community via a research-focused website or web page. The web is an effective tool for informing the research community about the agency's information needs, research protocol and administrative databases.

## ENHANCEMENT STRATEGY 9

### Develop user-friendly codebooks and data dictionaries.

#### Overview

In order to facilitate research partnerships and produce high-quality policy research, state agencies should inform data users of the content and condition of agency data. It is essential that researchers thoroughly understand the data set they will be using, particularly any limitations stemming from incomplete or unreliable data entry in certain fields. Tools like codebooks and data dictionaries that are easy to understand (e.g. acronyms defined, accessible language and terminology) can greatly assist in this process. To the extent possible, agencies should also share the policies and procedures that underlie the administrative functions that the database supports. This background information is very helpful in setting the context and limits of the data.

In addition to circulating useful tools, agency staff should be prepared to work with researchers to ensure sufficient understanding of the data. This can help researchers develop sound research designs and safeguard against erroneous data interpretation – both of which are beneficial outcomes for state agencies.

**In addition to circulating useful tools, agency staff should be prepared to work with researchers to ensure sufficient understanding of the data.**

#### Guidelines for Successful Implementation

1. *Integrate the following components into a data dictionary or codebook:*
  - Data origination source (i.e. mainframe from which data set was extracted)
  - Date of data collection
  - Description of data collection methods (e.g. phone survey)
  - Type of file (e.g. ASCII data file)
  - Contact person for additional information
  - Overall case count (i.e. number of records in data set)
  - Overall variable count (i.e. number of variables in data set)
  - List of variables
    - For each variable within data set:
      - Variable name and number
      - Short variable description, as well as any additional definitions or explanation of references needed to understand the meaning of the variable
      - Code format (e.g. numeric, text) and variable width (i.e. number of characters)
      - Variable location within data set (e.g. column 57)
      - Number of active records
      - Number of null or missing values

- Data collection materials (e.g. survey questions and response categories)
- Special notes or instructions, such as guidance on weighting data
- Restrictions on usage

### Tips and Special Considerations

- **Go electronic.** Post data dictionaries and codebooks online and/or make them available in electronic format so they are easier for researchers to access and navigate.

## ENHANCEMENT STRATEGY 10

### Develop formal data sharing agreements.

#### Overview

To standardize the research process and promote policy-relevant research, state agencies should each collaboratively develop a model Memorandum of Agreement (MOA) for data sharing. MOAs are formal documents of agreement among state agencies or between a state agency and researcher. When the agreement is entered into with external researchers, it should include the agency research protocols (see previous enhancement strategy).

MOAs generally specify the data to be shared, the procedures for accessing the data, the safeguards to assure privacy and confidentiality, the uses of the data, and the protocol for making the findings public. Overall, MOAs spell out the mutual responsibilities of all parties and in some cases, assign specific timelines for meeting those responsibilities. A well-designed MOA can help facilitate the research process for agency staff as well as outside researchers. In reality, MOAs both in Connecticut and in other states range from detailed and well designed to general and lacking specification.

#### Guidelines for Successful Implementation

1. *Integrate the following elements into a model MOA for a state agency or multiple agencies:*
  - **Project Contacts/Data Sharing Partners.** Contact information for key agency or organization representatives should be included.
  - **Project Purpose and Scope.** A brief description of the project should be provided as well as the specific objectives of the research. If the project involves the sharing of data between two agencies, the information needs of the parties should be specified.
  - **Period of Agreement.** The effective and termination dates for the Agreement should be specified.
  - **Description of Data.** In this section, the parties involved provide detailed information about the data to be shared or exchanged. This could include the specific files from which the extracts will be drawn, the years covered, or the data elements.
  - **Roles of Participating Parties.** The responsibilities that each party will assume for data sharing and utilization should be delineated. Deadlines for assigned tasks, such as extracting the dataset or delivering the draft to the agency for review, should be specified, where appropriate.
  - **Justification for Access.** In some circumstances, the access to certain data may be justified under federal or state legislation. Where appropriate, the researcher should cite the legislation under which the data request is being made.

- **Privacy and Confidentiality Assurances.** This section typically spells out the administrative, technical and physical safeguards that the researcher will take to protect the confidentiality of the data. The relevant federal and/ or state legislation and/or guidelines concerning privacy and confidentiality are cited that will govern the research project. It should also list any specific uses of client sensitive or confidential data (i.e., social security number, wage reporting) that will be prohibited under the MOA.
- **Repository of Data and Security Safeguards for Data Access.** The researcher should specify where and how the data would be stored, and the means for controlling unauthorized user access to it.
- **Method of Data Access/ Implementation.** This section should identify the means by which the requested data will be made available, e.g., electronic transmission over the Internet, hard copies of data extracts. It should also specify whether transmitted data would be encrypted.
- **Data Sharing Project Costs.** Any costs associated with data sharing should be identified as well as who will assume the responsibility for meeting them.
- **Uses of the Data.** The researcher should delineate the specific uses of the data and how they relate to the project's objectives. If data are being shared between two agencies, the parties should specify how the data would be used to meet their respective information needs. Where new integrated data sets are being created, the researcher should outline the specific steps taken to integrate the data. The data elements contained in the new data set should also be identified.
- **Agency Review of the Findings.** This section should identify the opportunities that the agency will have to review and to provide input on the findings before they are made available, publicly. Sufficient time for review should be provided. The agency's right to include comments in the report should be spelled out if that is an agency requirement.
- **Materials and Property/ Ownership of the Data.** Where the particular use of findings or reports requires the prior written approval of the state agency, it should be specified in this section. In addition, the ownership of the data sets upon the project's conclusion should be indicated.
- **Notice.** Specification should be made that any proposed changes in the MOA should be made in writing and addressed to the designated contact persons responsible for the research project.

***Example:** Data Sharing Agreement Between the Connecticut Departments of Public Health (DPH) and Social Services (DSS)*

As a result of Connecticut legislation enacted in May 2002, DPH and DSS entered into a Memorandum of Understanding to facilitate sharing of data, including

HUSKY (Medicaid) and birth data. Specifically, some of the goals of the MOU include increasing coordination of programs funded by the Maternal and Child Health Block Grant within each agency and reducing duplication of effort while allowing joint access to critical Medicaid and public health data. (For a copy of the MOU, see the Resources/References section).

## Tips and Special Considerations

- **Share the findings.** Once data has been shared and the analysis is complete, it is important to share the findings with the public in a timely manner and in an accessible way. Most people have come to rely on the Internet for information, thus state agencies and researchers should endeavor to share information, whenever possible, through the web. Published reports should include substantive information and should be presented in a user-friendly format.



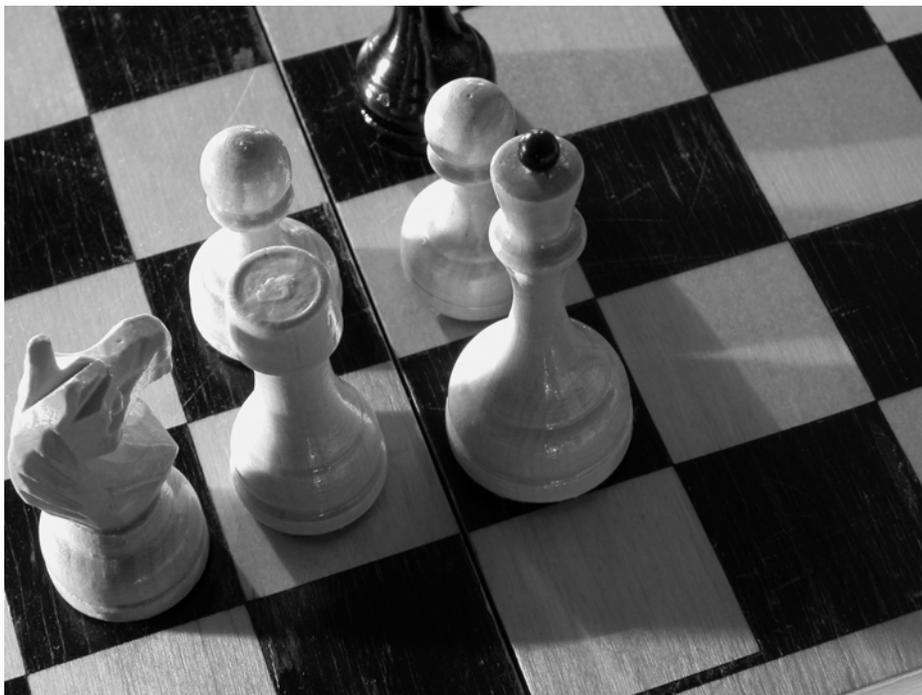
## Thinking Strategically: *Research Capacity Enhancements*

---

Research is, or should be, a valuable tool for planning and development within state agencies. Evaluation data, trend analysis and other research mechanisms can greatly assist agencies in identifying the need for new programs and constantly improving service delivery. For agencies to truly reap the benefits of research, it requires thoughtful planning. Agencies must consider what research questions or topics are most informative to their work and establish processes for gathering the data and facilitating such research. Given limited resources, state agencies often do not have the internal capacity to analyze administrative data for research, planning and development purposes. Therefore, developing adequate research capacity will often involve teaming up, formally or informally, with external researchers.

The guidelines in this section will walk users through four strategies for developing research capacity within state agencies. Although many of the strategies are directly applicable to facilitating collaborations with external researchers, overall these strategies represent incremental steps that state agencies can take to entrench research and evaluation within agency culture. These strategies are:

- Integrate evaluation/data collection into program design
- Develop a research agenda
- Create a research committee
- Develop research partnerships



# ENHANCEMENT STRATEGY 11

## Integrate evaluation/data collection into program design.

### Overview

When structuring a new service program, it is critical to think strategically about how to support program goals through the program's design. Part of this strategic planning should include incorporating evaluation components that will help measure progress toward program goals. Through program evaluation, administrators can identify areas in need of improvement and ultimately, enhance service delivery and strengthen the program. Evaluation data that documents positive program effects can also help secure continuation or expansion funding.

Program architects must also think strategically about how evaluation data will be collected so that the data can support a variety of evaluation and research functions. Although the primary purpose of the data collection may be to administer and track program inputs, opportunities to expand data utility should be pursued, where feasible. Such opportunities might include coding variables in a manner that allows for more detailed analysis (e.g. capturing specific ages rather than coding them into age ranges, like "under 18 years old.") Of course, data collection efforts should not be made too cumbersome or data quality may suffer (e.g. low response rates or incomplete data entry).

**Example:** *RIte Care Medicaid Managed Care Program (Rhode Island)*

RIte Care's database design reflects the program's long-term health care goals and provides a framework for evaluation of progress towards these goals. For example, contracts established with private health plans specify *quality and access standards* as well as *performance measures*, both of which are subject to ongoing monitoring by program staff. To meet these monitoring requirements and assess RIte Care's impact on service delivery and meeting unmet health care needs, RIte Care's Performance Goal Program and its Research and Evaluation Studies Program were established. These programs collect and analyze data and serve as a built-in bridge between practice and research, and ultimately policy.

### Guidelines for Successful Implementation

1. *Determine program goals.* This will help provide a framework for the evaluation.
2. *Decide what should be evaluated.* In addition to evaluating progress toward program goals, state agencies may want to add other components to the evaluation.
3. *Construct indicators or performance measures for each evaluation component.* This step in the process will help determine what data should be collected and in what format. The following performance measures/goals were developed by Rhode Island to assess health quality and access through their RIte Care program.

### **Administration/ Management**

- Permanent ID cards issued within 10 days
- Primary Care Providers assigned within 10 days
- Average speed to answer calls, 30 seconds or less
- Grievance and appeals in specified time frames
- Pay “clean” claims within 30 days
- Pay claims to ER for medical screening
- Notify Department of Human Services of potential third-party liability within 15 days

### **Access**

- Members seeking ER receive services immediately
- Members seeking urgent care receive services within 24 hours
- Members seeking non-ER, non-urgent behavioral health care within 5 days
- New adult members receive first PCP visit within 6 weeks
- New pediatric members receive first PCP visit within 90 days

### **Clinical Services**

- Members less than 2 years old are immunized
- Members between 6 and 21 years receive age appropriate screenings
- Pregnant women receive adequate prenatal care using Kotelchuk Index
- Lead screening for age 18 months received in preceding 9 months
- Specified Pap rates for women age 16 to 64 years

### **Pilot Performance Goals**

- Postpartum visit within 21 days to 56 days after delivery
- First outpatient pediatric visit for infants within 2 to 4 weeks
- ER visits for persons with asthma, using HEDIS measures for 4-9 year olds and 10-17 year olds
- Outpatient visit after discharge for a mental health diagnosis within 2 weeks
- Translation assistance offered by member’s doctor or health plan

### **Don’t Underestimate the Possibilities with Indicators**

Indicators are valuable tools for measuring progress toward not only program-specific goals, but also larger (and often interagency) goals such as reducing child poverty. For example, the national School Readiness Indicators Initiative was designed to help states track young children’s progress toward the important yet expansive goal of “school readiness.” Connecticut and other states participating in this initiative each developed a set of school readiness indicators using data from multiple agencies and issued reports to help states measure progress going forward. Connecticut’s report can be downloaded at [www.chdi.org](http://www.chdi.org) or [www.gettingready.org](http://www.gettingready.org).

4. *Consider other databases containing information that will enhance the evaluation and structure the new administrative database so that it easily links with those other databases.*
5. *Establish who will be responsible for data collection and analysis of that data.* The responsibility for primary data collection and/or analysis may be the state agency, an agency grantee or consultant, or a combination of both.

## Tips and Special Considerations

- **Consider a logic model.** A logic model may be a useful tool in outlining what the program goals are and how progress toward those goals will be evaluated.
- **Include researchers in the conversation.** When constructing indicators/performance measures and determining how data will be collected, it might be beneficial to invite researchers/evaluation experts to the table. Their guidance, particularly during the design phase, will help ensure sound evaluation data in the long run.
- **Proxies may be necessary.** Some program outcomes may be difficult or costly to measure reliably. In such cases, using data proxies that indirectly measure the desired outcome may be useful tools.
- **Share the learning.** Remember to share evaluation results with key stakeholders (e.g. service delivery staff, program funders) so these findings become an effective tool for accountability and, ultimately, program improvement. Also, consider sharing results from others states, where possible, to foster competition.

## ENHANCEMENT STRATEGY 12

### Develop a research agenda.

#### Overview

A research agenda outlines a state agency's research priorities and establishes a unified direction for research endeavors. When properly designed, it is a powerful strategic planning tool that leads state agencies toward continuous improvements in policies and practices. Developing a truly directive research agenda should be a collaborative process involving both those who have information needs and those who will conduct the research to meet those needs. Soliciting input from research universities, colleges and policy organizations will help determine what research will have the greatest impact on informing agency policies and practices.

Another benefit of opening up the agency's planning process in this way, is that it helps develop relationships between departmental staff and outside researchers. In building consensus around the research agenda, both parties can gain a better understanding of the motivations and constraints that influence each other. This sets a firm foundation for future public-private research partnerships (discussed later in this section).

#### ***Example 1:** Illinois Child Welfare Research Agenda*

In 1996, the Illinois Department of Children and Family Services (DCFS) and the Illinois School of Social Work launched the Children and Family Research Center. The Center's initial objective was to develop a strategic plan that would establish priorities for DCFS and Center research endeavors and set forth three- to five-year research goals. After an 18-month information gathering and consensus building effort, they succeeded in creating a unifying agency for statewide child welfare research.

#### ***Example 2:** Minnesota Child Care Policy Research Partnership*

The Partnership is a collaboration among several state agencies, counties, child care resource and referral agencies, and university researchers. Its research agenda is designed to answer critical questions about how affordability, quality and accessibility of child care affect outcomes for families and children. An important research objective is to inform policy-makers about the impact that subsidy levels, tiered reimbursement, quality regulations and standards, and other state policies have impacted child care quality.

#### Guidelines for Successful Implementation

1. *In addition to state agency staff, invite researchers and relevant content experts (e.g. child welfare, early care and education) to join the steering committee.*
2. *Identify key stakeholders (this will likely include more than just steering committee members) and begin gathering information on what they believe are the priority policy and practice areas and what specific research questions they would like to see incorporated into the research agenda.*

3. *Clarify and prioritize questions/issues generated from the information gathering process.*

***Example 1:*** *Connecticut Consortium for Applied Child Welfare Research*

The Consortium, situated within Connecticut's Department of Children and Families, evaluated each proposed research topic within an analytical framework that consisted of the following components:

- Prospective Research Questions – Detailed policy and practice questions related to the topic that could be the subject of in-depth inquiry
- Rationale for Research – Basis upon which DCF should support research initiatives in the topical area
- Benefits of Research – Specification of potential positive outcomes for policies, programs and practices at DCF
- Availability of DCF Resources for Undertaking Research – Discussion of the resources available at DCF and their adequacy for pursuing the intended research activities
- Role of Colleges and Universities in Undertaking Research in the Topical Area – Discussion of the role that colleges and universities may assume in pursuing research activities on the topic and the potential for DCF entering into a contract with an outside research institution

***Example 2:*** *Children and Family Research Center (Illinois)*

In developing their child welfare research agenda, the Center used The Delphi Technique to gain consensus on which research questions to include in the final agenda. Questions were categorized under the following service delivery headings: child protection; family maintenance; substitute care; family reunification; permanency planning; and substance-exposed infants.

4. *Outline a preliminary research agenda and circulate to key stakeholders for comment.*
5. *Disseminate the final research agenda within and outside of the agency. If the agency has an IRB, ensure that they are well versed in the research priorities outlined within the agenda.*
6. *Develop a strategic plan for research based on the agenda.*
7. *Establish a process for periodic stakeholder input and review of the research agenda.*

## Tips and Special Considerations

- **Make it relevant.** A general theme that emerged from states developing research agendas was that research priorities should be directed at service/intervention outcomes that can inform policy and practice. Also, research should examine the impact that existing policies have on client well-being and caseload dynamics.
- **Try different methods.** Various methods can be used to gather information from different stakeholders, including site visits, mailed surveys and focus groups. Using a variety of methods may help generate more diverse feedback.

- **Circulate the agenda.** To help circulate the research agenda to researchers outside of the agency, consider posting it online, perhaps on a dedicated research website or web page. Another option is emailing the agenda to key contacts in relevant university departments or other research organizations.

### **Taking the Research Agenda to a Higher Level: Interagency Collaboration**

Agencies can strengthen their research agendas (and the subsequent information they generate) by partnering with each other during the development phase. Any agency can take the initiative to collaboratively develop research questions or identify goals that are of mutual interest given each agency's objectives and service populations, and to integrate these elements into a research agenda. Bringing various agencies to the table for this purpose may be a one-time collaboration, or may be a function of a formal *interagency* research committee. Research committees are discussed in the next section.

## ENHANCEMENT STRATEGY 13

### Create a research committee.

#### Overview

A research committee functions as a coordinating body for all research activities occurring within (or across) state agencies. Although the core members are likely to be agency staff, it is also important to have strong representation from higher education institutions. While both public and private partners are often “at the table,” research committees are generally less formal and enduring arrangements than official research partnerships. The research committee can actually be the precursor to the research partnership, which is discussed in the next section.

Some of the key functions of a research committee include:

- Develop and implement an annual Research Agenda to guide the development and implementation of research activities
- Coordinate research activities with federal and state reporting or statutory requirements.
- Create a communication linkage between the state agency(s) and outside researchers to facilitate the conducting of research and to make existing data available for research purposes in a timely and consistent manner.
  - This might include maintaining a website with information on current research projects and links to key resources for those interested in conducting new research (such as the agency IRB).
- Foster coordination with the agency IRB (if one exists) to promote continuity between research inquiry, project design and implementation, and follow-up on recommendations.

#### Guidelines for Successful Implementation

1. *Establish the committee with a charge from agency leadership.* Though the full scope of the committee is likely to evolve over time, it is important to begin the work with an understanding of the goals that commissioners want the committee to achieve. The commissioner or his/her representative should convene the group for at least the first meeting. Frequent communication with the leadership of the agency or agencies should be maintained.
2. *Committee membership should include key stakeholders within the department(s).* A balance of information technology, research and program/policy staff is important. Sometimes there are other interested persons who can facilitate the work. For example, some agencies have very good facilitators in their training units who are particularly adept at helping a new task group establish itself.

In addition, the committee should have some representation from higher education faculty and other researchers. It is probably not necessary to have all the institutions at the table. Other researchers can be brought in when there are specific projects and studies.

3. *Develop a workplan.* The committee needs to have tasks and a realistic timeline, so that it does not become an advisory body. Development of a research agenda, review of databases, assessment of the agency's research infrastructure and holding research roundtables are examples of tasks that the committee could decide to do. The activities should be natural outgrowths of the goals the committee and agency leadership have established.
4. *Build in an evaluation of the committee's progress.* From the beginning, the committee should have a framework for evaluating its process and reporting back to agency leadership.
5. *Consider transforming the research committee into a more formal research partnership.* This strategy is discussed in the next section.

### Tips and Special Considerations

- **Committee structure can be flexible.** In Connecticut, the Department of Children and Families' research committee is structured as an advisory committee to a more formal public-private partnership – the Connecticut Children Welfare Research Consortium.

## ENHANCEMENT STRATEGY 14

### Develop research partnerships.

#### Overview

Research partnerships yield many benefits for state agencies and outside researchers alike. Given that these arrangements are more permanent and enduring, they provide more opportunities to foster trust among state agencies and researchers. Furthermore, they enable researchers to develop specific content knowledge about services, which, combined with their academic and technical expertise, results in higher quality research with greater practical application. Through these partnerships, state agencies can voice their needs and help researchers understand what information would help inform their policy and programming decisions. Likewise, researchers can help state agencies understand what barriers exist within the agencies' current data and research infrastructure that impede the type of policy-relevant research the agencies are seeking. Together, both parties can work to address these issues and increase the volume and value of research conducted with administrative data over time.

These permanent partnerships can help state agencies and researchers achieve several key goals:

- Identify and draw down federal funding and other grant opportunities
- Contribute toward improving the quality of services delivered to children and families
- Meet research and analysis needs for reporting requirements to funders, policymakers and others
- Attain academic goals for faculty (publishing) and students (curriculum/experience)
- Capitalize on federal reimbursement for research and training
- Assist state agencies to analyze administrative data
- Bring community partners to the partnership collaboration
- Implement a policy-relevant research agenda
- Apply research expertise to public policy issues
- Bring research on best practices and social issues to the attention of agency staff

#### ***Example 1:*** *Minnesota Child Care Policy Research Partnership*

The Partnership is a collaboration among several state agencies, counties, child care resource and referral agencies, and university researchers. The Partnership's primary goal is to foster sound research on child care issues of importance to policy-makers at the local, state and national levels. Funding is made available through a grant from the U.S. Department of Health and Human Services, Child Care Bureau. The Partnership identified a set of five inter-related research study areas designed to provide a comprehensive assessment of child care in Minnesota and its impact on children and families. They are:

- Quality of Care
- Parents' Choice of Type of Care
- Employment, Earnings and Job Stability of Parents Receiving Child Care Assistance
- Impact of Tired Reimbursement on the Market for Child Care

- Impact of Subsidies on the Market for Child Care

***Example 2: Children and Family Research Center (Illinois)***

In 1996, the Illinois Department of Children and Family Services (DCFS) and the University of Illinois School of Social Work launched the Center. The Center's primary functions are to conduct research and report on the outcomes of DCFS intervention and to evaluate service delivery and practice. Management of the Center is the responsibility of the DCFS Research Director. In this capacity, the Director oversees all research conducted with public child welfare clients and service delivery systems, staffs DCFS' IRB, and conducts and contracts for DCFS sponsored research. An advisory board composed of researchers, DCFS staff, child welfare agency representatives and child advocates provides oversight. DCFS provides the funding for the Center.

## Guidelines for Successful Implementation

1. *Clearly articulate the roles and expectations of the partners.* Involving researchers with different interests in a cooperative project requires a strong commitment to developing and maintaining collaborative relationships. State agencies operate in a very different world with substantial constraints and potential for extensive impact. It is therefore important to identify and clarify each partner's needs and expectations. A focus group could serve as an ideal methodology for uncovering these perspectives. Based on the findings, member roles and responsibilities should be spelled out and ratified.

Among the issues that the partnership will initially have to consider are the following:

- How will the partnership move beyond self-interest to a shared vision and common understandings?
  - How will decisions be made that incorporate different motivations and priorities?
  - How can communication be facilitated between members?
  - Are participating organizations committed to the partnership or is it limited to the participating member?
  - Who has the authority to speak for the participating organization?
  - Which decisions will be made by the partnership as a whole and which will be made by persons with expertise in different areas?
  - What will the partnership's products be?
2. *Charge the steering committee with developing recommendations on the partnership's purpose, functions and organizational framework.* Essentially this work group should be charged with the responsibility for determining how the initial gathering of departmental staff and outside researchers can be transformed into a permanent and meaningful partnership. Discussion should focus on mission and purpose, roles and expectations of partners, potential research activities, and protocols for carrying out the partnership's work.
  3. *Develop a research agenda to guide the prioritization and content of partnership research activities.* (see Enhancement Strategy 12 for more information on developing a research agenda).

4. *Foster a cooperative working relationship between the research partnership and the department's IRB (if one exists).* The IRB has the capacity to promulgate the partnership's research priorities and to achieve consistency in the research efforts supported under the department's aegis. Close ties with the IRB increase the likelihood that it will adopt the partnership's priorities as its guidelines for reviewing and approving research proposals it receives.

## Tips and Special Considerations

- **Consider the dynamics.** In determining the partners, consider whether it is most effective to work with one or two researchers, one department of a university, a multi-department center, or even several institutions of higher education.
- **Look beyond higher education institutions.** Consider including non-profits and consulting firms that have strong research capability.

## Endnotes

---

<sup>1</sup> Early Childhood DataCONNECTIONS is a public-private partnership of the State of Connecticut's Department of Social Services and the Child Health and Development Institute of Connecticut. The project's mission is to promote well-informed decisions on policies and programs for young children by improving state agencies' research capability.

<sup>2</sup> Given the project's emphasis, the scan was limited to databases collecting information on young child populations and services for children and families. Over 80 databases were examined.

<sup>3</sup> For more information on this initiative, visit [www.allkidscount.org](http://www.allkidscount.org).

<sup>4</sup> National Research Council and Institute of Medicine. (October 2004). *Children's Health, The Nation's Wealth: Assessing and Improving Child Health*. Report Brief. [http://books.nap.edu/html/child\\_health/reportbrief.pdf](http://books.nap.edu/html/child_health/reportbrief.pdf).

<sup>5</sup> Hotz, V.J, Goerge, R., Balzekas, J. & Margolin, F. (1999). *Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development*. Joint Center for Poverty Research, Northwestern University/University of Chicago. [www.jcpr.org/report.html](http://www.jcpr.org/report.html).

<sup>6</sup> Hotz, V.J. et. al. Ibid.

<sup>7</sup> For more information on this methodology, see *Modern Structured Analysis* at <http://www.yourdon.com/publications/index.html/>.

<sup>8</sup> National Institutes of Health Data Sharing Policy and Implementation Guidance is available at [http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm).

