2009

Connecticut Health Information Plan
A roadmap for improving access to health data

University of Connecticut
Center for Public Health and Health Policy

OFFICE of
HEALTH CARE
ACCESS
Executive Summary

Despite an enormous investment of effort and resources in data collection, Connecticut finds itself in the same position as many other states with respect to the difficulty of interchanging data among its disparate data systems. Data on the same individuals is often fragmented among multiple databases, both within and across agencies, and Connecticut lacks a technological and institutional platform through which this information can be easily integrated and shared. As a result, it can be time-consuming and labor-intensive to access the data required to devise more effective policies and programs to reduce morbidity and mortality in the state, better coordinate the delivery of services, evaluate best-practices, and monitor the health of our most vulnerable populations.

The Connecticut Health Information Network (CHIN) initiative constitutes a new paradigm for collaboration in information sharing. This initiative aims to create an innovative technological platform for accessing and integrating information from different data sources in a highly secure manner. It will establish an institutional framework to support government-university-community partnerships and collaborations around access to and analysis of health data. The initiative is organized by the Center for Public Health and Health Policy (CPHHP) at the University of Connecticut and is being developed by the CPHHP in collaboration with the Connecticut Departments of Public Health, Developmental Services, Children and Families, Information Technology, and the Office of Health Care Access.

In 2007 the Connecticut General Assembly passed Public Act 07-02, June Special Session, charging the University of Connecticut Health Center and the Department of Public Health with the development of a Connecticut Health Information plan. Specifically, UCHC and DPH were charged with (1) describing existing health and human services data; (2) developing an inventory of the various health and human services data aggregation initiatives currently underway; (3) including a framework and options for the implementation of a Connecticut Health Information Network, including query functionality to obtain aggregate data on key health indicators within the state; (4) identifying and proposing measures to comply with confidentiality, security and privacy standards; and (5) including a detailed cost estimate for implementation and potential sources of funding. Money was appropriated to the University of Connecticut Health Center for the continued development of the network and the CPHHP convened teams of representatives from the named agencies to explore the technological and policy issues set out in P.A. 07-02. The CPHHP is the principle author of the plan that we present in the attached report. It has been carefully constructed in collaboration with the agencies named in the CHIN bill and provides a framework for surmounting the technical, logistic, legal and financial challenges to integrating Connecticut’s fragmented health data.

With funding from the General Assembly for FY 2008 and 2009, the Connecticut Health Information Network has moved from the Prototype phase, funded by the University in 2005 and 2006, to a Limited Production deployment that has allowed the University to conduct a formal assessment of the cost and feasibility of deploying a fully functional network. The current phase of the initiative is still in process and is focused on evaluating, expanding, and refining the current Limited Production system to prepare for and assess the feasibility of a Full Production launch.

The proposed CHIN infrastructure utilizes a federated or distributed database architecture that provides record integration of multiple independent data sources in a secure networked environment. It allows access to the integrated data through a secure web-based query capability. This architecture will provide users with the benefits of a data warehouse while avoiding much of the costs and disruptions associated with implementing a physical data warehouse. The federated approach (which is also employed on the national level by CDC’s Public Health Information Network [PHIN] and National Electronic Disease Surveillance System [NEDSS] programs) facilitates data sharing by defining a common set of standards & protocols and providing middleware software components to enable messaging and data exchange among the data sources. The CHIN physical architecture consists of services running on servers at the University of Connecticut Health Center. These services include query functionality, which allows authorized users to submit customized queries that access data from multiple
data sources; the metadata registries, which provide information about the types of data stored in each data source; and the trusted broker, which executes the approved queries and performs the matching and de-identification functions while ensuring the security of the data. The data sources themselves remain in each agency and can be accessed by the mapping and transport software components of CHIN which operate independently from the data sources. This enables integration of data from the various sources through the CHIN service components at UCHC via encrypted connections with certificate-based authentication.

The network is highly secure and access is tightly controlled. Users must have authorization to access CHIN, and only de-identified data will be made available to them. The vast majority of CHIN’s application components are open-source (non-proprietary and freely available at no cost), enabling the state to avoid the costly and restrictive vendor agreements commonly associated with proprietary systems. The open source nature of CHIN will also provide opportunities to access research and development funding from federal sources such as the National Institute of Health, which has a stated preference for open source systems.

A comprehensive review of databases identified as candidates for inclusion in CHIN supports the potential utility of cross-agency data sharing and integration for research and evaluation. This report includes a thorough review of the other information technology initiatives currently underway in selected state agencies. A number of initiatives to improve data access have recently been completed or are in progress. These efforts focus primarily on increasing the accessibility of individual data systems within the host agencies, or on providing access to aggregated data for reporting and monitoring purposes. They are an important part of the state’s overall data quality and data access improvement strategies.

Most current efforts, however, do not seek to integrate data across different databases, particularly across different agencies. They are limited in scope and have not developed the technologies that will enable platform-independent integration across a range of data sources and across agencies. In this regard, CHIN’s potential contributions to these endeavors are unique. CHIN will complete the data improvement strategies of the agencies by providing the connectivity among databases and across agencies that will give true functionality to Connecticut’s “data system.”

A governance structure is proposed to guide and oversee the development and operation of CHIN. A governing body composed of representatives from the participating agencies would devise policies and procedures governing access to data through CHIN and would review all requests for access to information, with each agency maintaining exclusive control over its data. The University of Connecticut’s Center for Public Health and Health Policy would house the administrative core for the network and have responsibility for the development, implementation, and maintenance of the network infrastructure and provision of IT services, in collaboration with DOIT and the participating agencies.

This report includes an analysis of the potential legal barriers to CHIN. It recommends a straightforward method for surmounting the challenges posed by HIPAA, through the execution of Business Associate agreements between UCHC and the participating agencies. State restrictions on the release of personally identifiable information maintained by the agencies named in Public Act 07-02, (these statutes are listed in Appendix A) can be addressed via “notwithstanding” legislation that will leave the restrictions in place for their original purposes while making clear that it is permissible for the agencies to share individually identifiable information with CHIN for purposes of data matching and de-identification and subject to the requirements of federal privacy laws.

Future development and production implementation of CHIN will require additional funding for the Center for Public Health and Health Policy at the University of Connecticut Health Center. Additional funding for the Department of Information Technology and for the agencies that house their own databases will also be required over the same period for these purposes. Based on the successful demonstration of the milestones in the feasibility study to be completed by the end of FY09 (as set forth in Section III), the following recommendations should be considered.
Recommendations

1. We recommend production deployment of the CHIN infrastructure described above for the five agencies named in Public Act 07-02, in accordance with DOIT’s Systems Development Methodology and subject to negotiation of Memoranda of Agreement and, where applicable, Business Associate Agreements with the Center for Public Health and Health Policy at the University of Connecticut Health Center, which will serve as CHIN administrator. We also recommend expansion of this network to include other state agencies as funding and federal regulations permit.

2. Other multi-agency data integration efforts should be encouraged to collaborate with CHIN in developing their initiatives, and where appropriate, to capitalize on the state’s investment in the technologies developed for CHIN.

3. Each agency should issue guidelines for human subjects’ protection review of data accessed through CHIN. In addition, the CHIN collaborative should seek guidance from the Human Subjects Protection Office of the Department of Health and Human Services regarding the protocol for Institutional Review Board (IRB) review of data accessed through CHIN. Efforts to streamline the review process should be explored, perhaps including the creation of a uniform IRB application and the eventual development of a central IRB for CHIN applications.

4. The Legislature should enact legislation permitting participating agencies to disclose identified information to CHIN, notwithstanding existing state statutes to the contrary, and subject to the privacy and confidentiality requirements of HIPAA and other federal laws and regulations. This approach would leave existing statutes in force for all other purposes. Language from legislation proposed in 2007 is attached as Appendix D and is suggested as a model for such a provision.

5. The Legislature should establish in statute a CHIN Governing Board to establish policies and procedures for access to data through CHIN. Language from legislation proposed in 2007 is attached as Appendix D and is suggested as a model for such a provision.

6. The UCHC, through its Center for Public Health and Health Policy, should be established as the Administrator of CHIN, with the CHIN Main Application running on network servers located at the UCHC. CHIN Remote Modules should be established at DOIT for participating databases that are maintained by DOIT and at participating agencies who maintain their own databases.

Summary

The Connecticut Health Information Network offers an unprecedented opportunity to enhance the state’s capacity to serve the needs of its most vulnerable populations; to improve the planning, delivery and coordination of health services; and to dramatically improve the health of Connecticut’s citizens. Recent political and technological developments – the burgeoning awareness of the advantages of data integration among the leadership at key state agencies, the successful collaboration of 6 state agencies in producing the Connecticut Health Information plan, and technological innovations that have substantially reduced the cost of large scale data integration projects – have created an environment in which the promise of an integrated health information network can be realized in the State of Connecticut. Moving forward with the development and production deployment of CHIN will require additional resources for the CPHHP, DOIT and the participating agencies in the coming years, some of which may be obtained from federal grants or other external funding sources.
# Table of Contents

Executive Summary .................................................................................................................. ii
Recommendations ......................................................................................................................... iv
Summary ..................................................................................................................................... iv
I. Introduction .................................................................................................................................... 1
   The Need of Interagency Data Sharing ......................................................................................... 1
   Barriers to Agency Data Sharing .................................................................................................... 3
   Surmounting Barriers to Interagency Data Sharing through the Connecticut Health Information Network .... 4

II. Inventory of Agency Databases ................................................................................................... 5
   Department of Children and Families .............................................................................................. 5
   Department of Developmental Services ............................................................................................ 6
   Department of Public Health ............................................................................................................ 8
   Office of Health Care Access .......................................................................................................... 10
   University of Connecticut Health Center .......................................................................................... 11

III. A Solution to the Data Integration Needs of the State of Connecticut ........................................... 12
   Technical Framework for Implementation of CHIN ......................................................................... 13
   Privacy, Security, and Record Linkage for Secure Data Sharing in a Federated Environment .......... 13
   Record Linkage .................................................................................................................................. 13
   Data De-Identification ....................................................................................................................... 14
   Secure Communication ..................................................................................................................... 14
   Access Control ................................................................................................................................. 14
   CHIN Functionality .......................................................................................................................... 15
   Architecture ...................................................................................................................................... 16
   Access to Aggregate Data ................................................................................................................ 17
   Community Concensus ..................................................................................................................... 18
   Development of a web-based query system ..................................................................................... 18
   Funding ........................................................................................................................................... 18
   Current Status of the CHIN Initiative ............................................................................................... 19

IV. Inventory of Health-Related Data Aggregation and Data Access Improvement Efforts ................... 20
   Single agency efforts ......................................................................................................................... 20
   Department of Children and Families .............................................................................................. 20
   Department of Developmental Services ............................................................................................. 21
   Department of Public Health ............................................................................................................ 22
   Office of Health Care Access .......................................................................................................... 24
   Other State Agencies ....................................................................................................................... 24
   Multi-agency efforts .......................................................................................................................... 26
   Solutions offered by CHIN ............................................................................................................... 28

V. Privacy and Confidentiality .......................................................................................................... 28
Federal Laws........................................................................................................................................29
Health Insurance Portability and Accountability Act (HIPAA), 42 U.S.C. 201 ................................29
Protection for Human Subjects 45 C.F.R. Part 46........................................................................29
Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. 1232g et seq., 34 CFR part 99. 30
Medicare/Medicaid Intermediate Care Facilities for the Mentally Retarded (ICF/MR), 42 CFR part 483.410 .30
State Laws .......................................................................................................................................30
Laws of General Applicability ........................................................................................................30
Agency Specific Laws, Regulations and Policies ........................................................................30
Department of Developmental Services ......................................................................................31
Department of Public Health .........................................................................................................31
Office of Health Care Access ........................................................................................................32
University of Connecticut Health Center .....................................................................................32

VI. Governance Plan Recommendations .....................................................................................33
Structure of CHIN ..........................................................................................................................33
Governing Board ...........................................................................................................................33
Administrative Core ......................................................................................................................33
Data Access Procedures ................................................................................................................34

VII. Costs and Funding ................................................................................................................34

References ......................................................................................................................................35
Appendix A: Substitute Bill 7376 (Effective 10/01/2007) .................................................................38
Appendix B: CHIN Work Group Membership ..........................................................................40
Appendix C: Federal and State Confidentiality Laws and Regulations ......................................41
Appendix D: Substitute Bill 7376 (proposed but not enacted) ......................................................42

Tables
Table 1. Health and Human Services Data: DCF ............................................................................5
Table 2. Health and Human Services Data: DDS ..........................................................................7
Table 3. Health and Human Services Data: DPH .........................................................................9
Table 4. Health and Human Services Data: OHCA ...................................................................10
Table 5. Health and Human Services Data: UCHC ..................................................................11
Table 6. DCF Data Access Improvement/Aggregation Efforts ......................................................21
Table 7. DDS Data Access Improvement/Aggregation Efforts ......................................................21
Table 8. DPH Data Access Improvement/Aggregation Efforts ......................................................22
Table 9. OCHA Data Access Improvement/Aggregation Efforts ...................................................24
Table 10. Other State Agency Data Access Improvement/Aggregation Efforts ............................25
Table 11. Multi-Agency Data Access Improvement/Aggregation Efforts ...................................26
I. Introduction

Connecticut has invested vast resources over the past few decades in collecting and maintaining data relevant to the health and well-being of state residents. These data currently reside in separate state agencies, each having an independent mission, governance, and technological infrastructure. For example, health surveillance data (i.e., immunization registry, lead screening), healthcare utilization data, documentation of child custody and support services, and data on developmental outcomes and academic achievement reside in the Departments of Public Health, Social Services, Children and Families, and Education, respectively. Moreover, key surveillance and administrative data that currently reside in the same state agency – e.g., immunization registry and death records – are maintained in separate databases which have extremely limited interoperability and do not readily allow for integrated statistical analysis.

Despite this enormous investment of effort and resources in data collection, Connecticut finds itself in the same position as many other states with respect to the difficulty of exchanging data among its disparate data systems. Information on the same individuals is often fragmented among multiple databases, both within and across agencies, and the state lacks a technological and institutional platform through which these data can be easily integrated and shared. As a result, it can be time-consuming and labor-intensive to access the data required to devise more effective policies and programs to reduce morbidity and mortality in the state, better coordinate the delivery of services, evaluate best-practices, and monitor the health of our most vulnerable populations.

Policymakers in Connecticut have had a longstanding interest in augmenting efforts to collect, analyze, and disseminate health information collected by the state’s executive branch agencies. In the past this desire was hampered by the technological challenges, political barriers, and financial burden accompanying a major data sharing and integration initiative. However, a burgeoning awareness of the need for data integration among the leadership at key state agencies, coupled with technological innovations that substantially reduce the cost of large scale data integration projects, have removed many of the most serious barriers to such a project.

In June 2007 the Connecticut General Assembly passed Public Act 07-02, June Special Session, charging the University of Connecticut Health Center (UCHC) and the Department of Public Health (DPH) with the development of a Connecticut Health Information plan of implementation. Specifically, UCHC and DPH were charged with (1) describing existing health and human services data; (2) developing an inventory of the various health and human services data aggregation initiatives currently underway; (3) including a framework and options for the implementation of a Connecticut Health Information Network, including query functionality to obtain aggregate data on key health indicators within the state; (4) identifying and proposing measures to comply with confidentiality, security and privacy standards; and (5) including a detailed cost estimate for implementation and potential sources of funding. The plan that we present below has been carefully constructed in collaboration with the agencies named in the CHIN bill and provides a framework for surmounting the technical, logistic, legal and financial challenges to such an ambitious data integration initiative.

The Need for Interagency Data Sharing

Connecticut’s repositories of health and human services data include many disparate information sources. As a result of the diversity of information systems, there is tremendous fragmentation of the data maintained on individual persons and families among the records in numerous state agency databases, with data on the same individuals spread across multiple systems. For example, youth served by the Departments of Developmental Services (DDS) and Children and Families (DCF) appear in the State Department of Education’s Public Schools Information System, in the Department of Public Health’s surveillance registries for immunizations and lead exposure, and may receive Medicaid benefits through the Department of Social Services (DSS). In a recent report prepared in response to a mandate from the Connecticut General Assembly, the Department of Mental Health and Addiction Services estimated that 60% of Connecticut residents receiving services from community mental health programs were on the caseload of at least one of eleven other state and federal programs, with
more than half of these (55%) also participating in benefit programs in the 2005 fiscal year.\textsuperscript{2} In another example, the University of Connecticut Health Center’s Neonatal Infant Systems Database tracks all infants admitted to the Neonatal Intensive Care Unit (NICU) at John Dempsey Hospital. These children frequently require services through the Birth-to-Three program administered by the Department of Developmental Services prior to moving into the educational system. Lastly, the University of Connecticut Health Center’s Quick Recovery dental record system has substantial overlap with multiple agency databases due to the Health Center’s role in providing dental care to approximately 60% of all Husky children and a sizeable percentage of Department of Developmental Services’ clients.

In order to develop optimal programming for the populations served by the programs of these Departments or to effectively evaluate those programs, it is often necessary to obtain data beyond that which is contained in an agency’s own databases, and to match records from the same individual across data sources. Examples of this need include DMHAS access to the DCF records of clients who are transitioning from DCF care to DMHAS’ adult mental health system to provide appropriate programs and services for this group. In addition, an accurate evaluation of DCF behavioral health services cannot be determined without the ability to follow this group as adults. Many DCF wards also transition to DDS services and supports as they age out of eligibility for DCF services. Again, DDS needs access to DCF records on this group, and DCF needs access to DDS records to fully evaluate the success of DCF programming for them. In addition, DPH frequently needs to integrate existing datasets for research projects, and to work collaboratively with State agencies that continually provide programmatic, socio-demographic, health status, and health service information that supports many important and vital policy developments. Epidemiological studies based on DPH data would also profit greatly from easy interoperability with OHCA inpatient discharge and outpatient encounter databases, as well as the other data systems maintained by OHCA.

However, the various data systems maintained by Connecticut’s state agencies have extremely limited interoperability, even within the same agency, which has severely limited intra- and interagency data integration and sharing. Current mechanisms for performing cross platform data integration are highly inefficient, labor intensive, and financially costly, severely limiting the availability of this wealth of data for research. Although not collected with research purposes in mind, administrative data such as this constitutes an indispensable resource for program monitoring, outcome evaluation, and efforts to establish best practices. Administrative data have many strengths: such databases tend to include large numbers of individuals, allowing for the study of relatively rare conditions and small populations; record-keeping tends to cover relatively long periods of time, offering the potential for longitudinal patterns to emerge; and the cost of accessing this information is trivial relative to the costs of obtaining comparable data through surveys or observational studies.\textsuperscript{3} Coupled with statistical information collected by states for surveillance and assurance purposes, administrative data offers an invaluable resource for improving public health. However, it can be extremely cumbersome to navigate among Connecticut’s various health databases for research purposes. Applications must be made to each agency’s IRB. Submissions to each agency and integration of data from different databases are laborious and time consuming.

Connecticut agency databases contain some very high quality data, and individual databases have supported significant research in the past. A review of published literature provides ample evidence that several agency databases contain high quality data capable of meeting stringent peer review standards. For example, in the past 15 years dozens of peer-reviewed articles using the Connecticut Tumor Registry have been published, many of them in the premiere medical and public health journals (e.g., JAMA, New England Journal of Medicine, Cancer).\textsuperscript{4} Analyses of data from the CT DPH’s adult lead surveillance database have been published in the Centers for Disease Control’s Morbidity and Mortality Weekly Report.\textsuperscript{5} Moreover, data from the DSS Medicaid enrollment and encounter database has appeared in numerous widely disseminated reports and peer-reviewed journals.\textsuperscript{6}
Barriers to Agency Data Sharing

While integrated information systems offer tremendous potential value to policy makers, agency personnel, and researchers, this potential is rarely realized due to the costliness along with political and technical difficulties associated with data integration and the repurposing of databases for health-related research. This is precisely the situation in which Connecticut finds itself: as is described in this report, complementary and overlapping information is stored in numerous discrete databases that are controlled by and physically housed in a host of different agencies, and are stored in disparate formats using a variety of hardware and software platforms with few or no universal identifiers to facilitate linkage across databases. This situation poses enormous barriers to epidemiologic research attempting to provide a more comprehensive profile of population health, as well as efforts to evaluate the effectiveness of particular policies and programs, both of which generally require information from more than one data source.

There are numerous technical challenges faced in attempting to integrate disparate health and human services databases.⁷⁸ An inventory of Connecticut’s child health databases completed by DataCONNections, a non-profit group associated with the Child Health and Development Institute of Connecticut, indicated that any effort to integrate Connecticut’s agency databases will have to resolve a series of technical and institutional challenges.⁹ A needs assessment conducted in 2005 by the University of Connecticut’s Center for Public Health and Health Policy amplified and expanded these concerns, identifying the following issues as barriers to the success of any statewide data integration initiative:

- **Security and Privacy**
  Access to personal health information (PHI) or other sensitive data must be carefully controlled and managed. Use of such information must adhere to a clear set of protocols consistent with agency guidelines, as well as federal and state laws.

- **Unique identification of individuals**
  Cases (e.g., individuals or families) do not have consistent, unique identifiers across the myriad of database systems across agencies, or even within agencies.

- **Accuracy and reliability of data**
  Data and variable names are not uniformly coded, data may not be recorded to the level of specificity desired by researchers, and may be ambiguous. Also, data quality may be less than ideal due to transcription errors or lack of proper archiving.

- **Legacy systems**
  Agency databases and computer systems are mainly designed for administration, case management, reporting, and accountability, not research. Additionally, these systems may be based on antiquated technologies that have not been updated for years or decades.

- **Disruption or alteration of business practices**
  Mechanisms necessary for access to agency information must be minimally disruptive, particularly in a climate of scarce resources.

- **Diverse organizational and institutional cultures**
  Different agencies may have vastly different institutional cultures and organizational practices, particularly with respect to access to information and the uses of that information in research. Compatibility in policies and procedures governing data release must be ensured.

- **Oversight**
  A governance system must be established that will facilitate access to data while protecting privacy and confidentiality and ensuring that state and federal laws limiting or constraining the uses of particular datasets are observed.
Surmounting Barriers to Interagency Data Sharing through the Connecticut Health Information Network

The Connecticut Health Information Network has emerged as an innovative and cost-efficient solution for improving access to the state’s health-related data. Launched as an initiative of the Center for Public Health and Health Policy at the University of Connecticut in 2004, CHIN currently exists as a working collaboration involving the University of Connecticut Health Center, five of the state’s executive branch agencies -- the Department of Developmental Services (DDS), the Department of Children and Families (DCF), the Department of Information Technology (DOIT), the Department of Public Health (DPH), and the Office of Health Care Access (OHCA) -- and faculty at the University of Connecticut’s Department of Statistics and Department of Computer Science and Engineering. Akaza Research, Inc., a medical informatics consulting firm in Cambridge, MA, has provided technological expertise to this project as a subcontractor of UCHC.

CHIN creates an infrastructure and environment that allows real-time access to integrated health information across state agency databases for policy makers, government officials and researchers. CHIN is designed to have a minimum impact on business processes at state agencies while providing several critical benefits to the state including:

• The ability to comprehensively assess target populations and their interactions with various state agencies over extended periods of time.
• Advancement of evidence-based practices through a more comprehensive evaluation of the link between programs and health outcomes over time.
• A tool to improve planning, service delivery and coordination, and budgeting for programs and services.
• Potential for knowledge transfer and the establishment of best practices for data collection and data management.
• A vehicle for state agencies to articulate their needs and agendas to the legislative and executive branches of government.
• The collaborative development of data standards resulting in greater consistency in the structure and content of databases within and across state agencies.
• The linking of data contained in legacy databases to data contained in current data systems within the same agency.

With funding from the Connecticut General Assembly in the 2008-2009 biennial budget, the University and its agency partners have been engaged for the past year and a half in developing a Connecticut Health Information plan. This plan is intended to establish a framework for developing and implementing a production network linking complementary and overlapping databases across agencies, to recommend legislative changes to facilitate and enable the integration of and access to data across state agencies, and to craft a governance structure to develop and administer protocols, rules, and procedures governing access to data. The plan includes the following key elements:

• An inventory of key individual-level databases maintained by agencies named in Public Act 07-02
• A technical framework for the Connecticut Health Information Network
• An inventory of other data integration initiatives in the State of Connecticut
• A review of state and federal laws governing data and data access in the 5 participating agencies
• A proposal for a governance framework that will promote access to data while simultaneously ensuring responsible use and compliance with all state and federal laws
II. Inventory of Agency Databases

Agency databases are maintained in several locations. DCF and DDS databases are located at and maintained by DOIT. DPH and OHCA databases are located at and maintained by each respective agency. For purposes of the development and implementation of CHIN, the location of a particular database is irrelevant. Both DOIT-maintained and agency-maintained databases can be included in CHIN.

Department of Children and Families

The Department of Children and Families (DCF) has several databases that contain data on DCF clients, facilities and providers. These databases include data on child neglect/abuse, DCF facilities and providers, juvenile delinquency, behavioral health and substance abuse by DCF’s clients. DCF’s principal child welfare database is LINK, a case management system which collects abuse and neglect data as well as client demographic data. LINK is a complex relational database which runs on an IBM DB2 database running on a Microsoft SQL Server. It contains over 200,000 records from 1996 to the present. LINK contains identifiers such as name, Social Security Number (SSN), date of birth, and ZIP code. A data dictionary does exist for LINK, but it is currently incomplete and in development. LINK was preceded by a legacy case reporting system to which it maintains a connection. DCF also has separate reporting and analytic databases that hold data transformed from LINK such as the University of Chicago’s Chapin Hall databases for foster care, and a web-based performance and outcome reporting system used for exit planning.

CONDoIT is DCF’s juvenile delinquency database and is a web-based application that runs on a Microsoft SQL server. DCF has a substance abuse database maintained by an IT contractor, Advanced Behavioral Health. This database contains results from the GAIN survey which focuses on behavioral health and substance abuse and is administered to most clients every two years. Lastly, many functional units have their own databases running on Microsoft Access that are used for their own business needs.10

Table 1. Health and Human Services Data: DCF

<table>
<thead>
<tr>
<th>Database</th>
<th>Platform</th>
<th>Size</th>
<th>Timeframe</th>
<th>Identifiers</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINK*</td>
<td>DB2/SQL Server</td>
<td>1.2 m individuals,</td>
<td>1996 - present</td>
<td>Name, DOB, Zip, SSN</td>
<td>Over 20 tables in database. Has links to legacy CMS and analytic and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>over 200k cases</td>
<td></td>
<td></td>
<td>reporting databases.</td>
</tr>
<tr>
<td>CONDoIT</td>
<td>IBM UDB to SQL</td>
<td>Over 200k individuals</td>
<td>2003 - present</td>
<td>Name, DOB</td>
<td>Data on children in Juvenile Justice centers.</td>
</tr>
<tr>
<td></td>
<td>Server</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KidPix</td>
<td>SQL Server</td>
<td>6k individuals</td>
<td>2000 - present</td>
<td>Demographic</td>
<td>Recent Images of children served by DCF.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>information from LINK</td>
<td></td>
</tr>
<tr>
<td>Reports_CT</td>
<td>SQL Server</td>
<td>1.2 mil</td>
<td>2005 - present</td>
<td>Name, DOB, Zip</td>
<td>Supports the Results Oriented Management System. Used to provide reports</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>on outcome measures required by the Juan F. Decree.</td>
</tr>
<tr>
<td>Chapin Hall</td>
<td>SQL Server/SAS</td>
<td>15 tables of child</td>
<td>2005 - present</td>
<td>No personal</td>
<td>Not an individual level database. Used for national-level comparisons of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>welfare-related data.</td>
<td></td>
<td>identifiers</td>
<td>child welfare.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Results Oriented</td>
<td>SQL Server</td>
<td>1.2 m individuals</td>
<td>2006- present</td>
<td>Unique person ID</td>
<td>This third-party system developed by Kansas University measures</td>
</tr>
<tr>
<td>Management System</td>
<td></td>
<td></td>
<td></td>
<td>ID number, Name,</td>
<td>numerous aspects of child welfare management and makes the web-based</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SSN, DOB</td>
<td>data available to Social Workers state-wide. The primary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The purpose of the system is to track measures that have been mandated by the Juan F decree, namely workload, visitation, and outcome aspects of child welfare management. The data are aggregated from the LINK Mirror database.

<table>
<thead>
<tr>
<th>Program and Services Data Collection and Reporting System</th>
<th>To be determined</th>
<th>To be implemented July, 2009</th>
<th>Name, DOB, Gender</th>
<th>Grant-funded external providers/programs for multiple DCF mandates</th>
</tr>
</thead>
</table>

*Targeted for initial inclusion in CHIN

**Department of Developmental Services**

The Department of Developmental Services (DDS) maintains several databases containing information on Connecticut residents with certain developmental disabilities. Their databases include information on the demographic characteristics of clients; needs assessments, including some health and diagnostic data; financial information; and quality review data including incident reports and their follow-up. DDS assigns a unique identification number to each client in its databases.

Most prominent among DDS individual-level databases are eCAMRIS, Level of Need (LON), and Planning Resources and Allocation Teams (PRAT). eCAMRIS is DDS’s consumer, demographic and provider/program information system which runs on SQL Server and has data on over 31,000 individuals since 1989, including over 15,000 currently active DDS clients. Data concerning client information, the services they receive, and their placements are maintained in a multiple-table structure. eCAMRIS contains identifiers such as name, SSN, date of birth, and ZIP code. Some historical data are kept concerning placements, providers and quality assurance.

PRAT runs on Microsoft Access and contains resource allocation plans for over 10,000 DDS clients dating back to 1996. PRAT has 3 main tables and can be integrated with eCAMRIS, which is used to populate the demographic fields of PRAT. However, PRAT has additional fields that are not present in eCAMRIS, such as funding allocations and dates of placement.

LON runs on SQL Server and holds data on level of need assessments for over 11,000 DDS clients since 2006. The data in LON include results from a survey containing over 100 questions that are used to determine services and placements for DDS clients, as well as some health and diagnostic information. LON is also linked to eCAMRIS and to DDS clients’ individual allocations of services.

DDS also maintains several other information databases including the Birth-to-Three and some Quality Management databases which run on Access/SQL Server. Birth-to-Three tracks children at risk for developmental delay or impairment through age three. Each entry in the various databases has a unique DDS identifier and includes other identifying information such as first and last name, date of birth and social security number. DDS has well-documented data dictionaries for these systems. There are some quality management systems maintained as Access programs that work with eCAMRIS data. These track aspects of licensing, abuse/neglect and mortality tracking.
Lastly, DDS will soon deploy an Individual Plan Budget database which will be used to collect and store data on associations of costs with clients and programs and supports. It currently is written in VB/SQL but soon will be rewritten as .NET/SQL for web-based access. The Individual Plan Budget database will take over some of the functions of eCAMRIS. 11

**Table 2. Health and Human Services Data: DDS**

<table>
<thead>
<tr>
<th>Database</th>
<th>Platform</th>
<th>Size</th>
<th>Timeframe</th>
<th>Identifiers</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>eCAMRIS*</td>
<td>.NET/SQL Server</td>
<td>Over 31k individuals</td>
<td>1989 - present</td>
<td>Name, SSN, Zip, DOB, DDS ID</td>
<td>consumer, demographic and provider/program information system.</td>
</tr>
<tr>
<td>Level of Need*</td>
<td>.NET/SQL Server</td>
<td>Over 11k individuals</td>
<td>2006 - present</td>
<td>Name, DDS ID</td>
<td>Contains data on level of need assessment for DDS clients</td>
</tr>
<tr>
<td>PRAT</td>
<td>VBA/Access</td>
<td>Over 10k individuals</td>
<td>1996 - present</td>
<td>Name, Zip, DDS ID</td>
<td>3 primary tables used for planning resource allocation for DDS clients</td>
</tr>
<tr>
<td>Quality Services Review</td>
<td>C#/SQL Server</td>
<td>New Database</td>
<td>2008 - present</td>
<td>Name, ZIP, DDS ID</td>
<td>Contains data on quality of services reviews, including licensing, incidents and follow up. Measures service outcomes and consumer satisfaction</td>
</tr>
<tr>
<td>Individual Budget Plan</td>
<td>VB/SQL Server</td>
<td>New Database</td>
<td>2008 - present</td>
<td>Name, DDS ID</td>
<td>Associations of costs with clients and programs and supports. Plan to be rewritten in .NET/SQL in approximately 1 year</td>
</tr>
<tr>
<td>Birth to 3</td>
<td>VB/SQL Server</td>
<td>15k individuals</td>
<td>1997 - present</td>
<td>Name, DOB, SSN, Parents’ names, DDS ID</td>
<td>Detailed data dictionary available</td>
</tr>
<tr>
<td>Web Res/Day</td>
<td>.NET/SQL Server</td>
<td>Approx 8,000 individuals</td>
<td>2008</td>
<td>Name, Program, Provider</td>
<td>Tracks billing for services based on eCAMRIS information</td>
</tr>
</tbody>
</table>

*Targeted for initial inclusion in CHIN*
**Department of Public Health**

The Department of Health (DPH) maintains over 300 diverse data systems ranging from vital records on Connecticut residents to disease-related screening and monitoring, licensing of facilities and providers, and program administration. These databases run on heterogeneous platforms. A listing of the main databases most relevant to CHIN is provided in Table 1 below and includes births, deaths, newborn screening, child health data and disease incidence.

The Birth Records database runs on an Oracle database platform and contains data on all Connecticut birth records since 1988 (records for previous years back to 1959 are stored in ASCII files). Identifiers in this database include names, dates of birth, social security numbers, and addresses for the infant and parents. However, social security numbers for the parents are not always available or reliable. Additionally, a state file number is assigned to each entry and is used as a unique identifier. This identifier is automatically generated and it is a concatenation of the year, state details and an auto-incrementing number. Currently, there are about 2.27 million births recorded in the database, with 40-45,000 new births added annually. The unique identifier from Birth Records is referenced in the Newborn Genetic Screening, Newborn Hearing, and Birth Defects databases and, in some cases, the Lead database.

Records collected through the EVRS Birth Records System are connected to several other DPH databases via data imports, record matches and/or references to birth certificate ID. These include Newborn Genetic Screening, Newborn hearing, Birth Defects, Immunization and Lead databases, all of which run on Oracle. The three databases concerning newborn screening and conditions date back to 2000 and contain over 350,000 records each. The Lead database dates back to 1994 and contains screening records on 730,000 individuals, while the Immunization Registry dates back to 1994 and contains records for 90,000 individuals.

DPH’s Mortality Records System holds data dating back to 1978 of all deaths that occur in the state of Connecticut. About 30,000 deaths are entered annually and each record includes ICD 9 or ICD 10 codes for cause of death. ICD 9 and 10 codes are computed by a third-party system from variables entered into the record. Each record includes a death record identifier as well as name and ZIP code. Records entered since 2005 are held in an Access program while records prior to 2005 are contained in a mainframe system.

The Women, Infants and Children database (WIC) is run on a mainframe system with a Foxpro interface. The WIC database holds information for over 300,000 WIC participants from 2003 to the present. Identifiers included in WIC are participant ID and name. Information on WIC vendors and associated financial data are also included in the database and some of these data predate 2003.

DPH’s Tumor registry contains data on reportable tumors in Connecticut residents. The Tumor Registry dates from 1973 to present and contains over 500,000 records. It runs on a platform consisting of Oracle, Java, XML and Linux OS. Identifiers held in the registry include patient ID, date of birth, SSN, ZIP, and, when applicable, Department of Motor Vehicles license number and date of death. Data from this Registry are reported annually to the National Cancer Institute\(^\text{12}\).
<table>
<thead>
<tr>
<th>Database</th>
<th>Platform</th>
<th>Size</th>
<th>Timeframe</th>
<th>Identifiers</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVRS Birth Records System*</td>
<td>Oracle, Sun Solaris OS Power Builder, MS-Access</td>
<td>2.27m</td>
<td>Records available 1959-present</td>
<td>Birth Certificate ID, ZIP, Names, DOB, SSN of infant and parents</td>
<td>Main birth records Collection system for births in the state of CT. Recently migrated records pre-2002 are under QA by Vital Records.</td>
</tr>
<tr>
<td>Mortality Records System</td>
<td>Access Path: Pre 2005: Mainframe fixed width w/FTP, COBOL, and/or EasyTree</td>
<td>30k annually</td>
<td>1978 - present</td>
<td>Death Record Identifier: YYYY-12345-SSN, ZIP, Name</td>
<td>Death records for state of CT. ICD 9, 10 codes computed for cause of death by third-party program.</td>
</tr>
<tr>
<td>Newborn Genetic Screening</td>
<td>Oracle9i, Access</td>
<td>Over 350k records</td>
<td>2000 - present</td>
<td>Accession number (8 digit number assigned at hospital). Also References Birth Certificate ID</td>
<td>Linked to EVRS. Record match routine with EVRS executed daily.</td>
</tr>
<tr>
<td>Newborn Hearing</td>
<td>Oracle</td>
<td>350k</td>
<td>2000-present</td>
<td>Main identifier is an automatically generated number assigned at the hospital. Also References Birth Certificate ID</td>
<td>Uses matching algorithms in Java and/or stored procedures to look up Birth Cert. ID from Vital Records.</td>
</tr>
<tr>
<td>Birth Defects</td>
<td>Oracle9i, Java Applets</td>
<td>Over 350k records</td>
<td>2000 - present</td>
<td>Main identifier is an automatically generated number assigned at the hospital. Also References Birth Certificate ID</td>
<td>In migration to new vendor w/Maven software. Expected implementation date is June, 2009</td>
</tr>
<tr>
<td>Immunization</td>
<td>Oracle(new) Foxpro (legacy)</td>
<td>Approximately 90k per month</td>
<td>1993 - present (also just over 1000k records prior to 1993)</td>
<td>State file#, Medicaid#, WIC#</td>
<td>Interfaces with Vital Records. Uses flat file export from EVRS. Interfaces for Medicaid MCO membership data for children ages 0-6.</td>
</tr>
<tr>
<td>Lead</td>
<td>Oracle 9i Sun Solaris OS</td>
<td>730k Individuals</td>
<td>1994 - present</td>
<td>Unique Patient Identifier assigned by the system</td>
<td>Populated by import of lab test results. Vital and Medicaid data are imported into the system twice yearly.</td>
</tr>
<tr>
<td>Women, Infants and Children</td>
<td>Mainframe: DB2, CICS, COBOL, FOCUS, JCL PC Server: Dos-Foxpro, Visual FoxPro</td>
<td>Over 300k participants</td>
<td>2003 - present for participants, goes back further for vendors and check records</td>
<td>Participant ID, Name</td>
<td>Mainframe and FoxPro hybrid platform</td>
</tr>
<tr>
<td>Tumor Registry</td>
<td>Oracle 10f, Java, XML, Linux OS</td>
<td>Over 500k</td>
<td>1973 - present</td>
<td>Patient ID, DOB, DOD, SSN, DMV license, ZIP</td>
<td>Reportable tumors in CT residents. Reported annually to the National Cancer Institute.</td>
</tr>
</tbody>
</table>

*Targeted for initial inclusion in CHIN
**Office of Health Care Access**

The Office of Health Care Access (OHCA) maintains seven databases containing data on the delivery of medical care and financial information from Connecticut’s hospitals. Four of these contain patient level data, while three contain only aggregate level data. The Acute Care Hospital Inpatient Discharge Database contains patient-level data obtained from discharge records collected from all 30 of Connecticut’s acute care hospitals from 1991 to the present. The fields include patient demographics, diagnoses, procedures, and charges, but do not include names; OHCA receives these data semi-annually from the Connecticut Hospital Association/CHIME with names removed. The database is run on SQL Server with an up-to-date dictionary of data definitions and contains over 6,900,000 records of discharges.

Other patient-level databases include the Outpatient Facility Encounter Level Database, the Emergency Department Database, and the Cardiac Registry. All three databases contain patient care data while the Outpatient Facility Encounter and Emergency Department databases also contain financial data associated with that care. The Outpatient Facility Encounter database runs on SQL Server and includes over 340,000 records related to patient care at licensed outpatient surgical centers required to submit the information as a part of their Certificate of Need approval. The Emergency Department Database is a flat file that contains patient care data from Connecticut’s acute care hospital emergency departments. The Cardiac Registry is an SPSS file that contains clinical data from 2004 to the present from seven acute care hospital cardiac programs that were established in the last five years. The cardiac registry holds approximately 17,000 records.¹³

**Table 4. Health and Human Services Data: OHCA**

<table>
<thead>
<tr>
<th>Database</th>
<th>Platform</th>
<th>Size</th>
<th>Timeframe</th>
<th>Identifiers</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Care Hospital Inpatient Discharge Database*</td>
<td>SQL</td>
<td>6.9m records</td>
<td>1991 – present</td>
<td>DOB, ZIP</td>
<td>Patient and financial data from all discharges from CT-licensed acute care hospitals</td>
</tr>
<tr>
<td>Outpatient Facility Encounter Level Database (CON)</td>
<td>SQL</td>
<td>Over 340k records</td>
<td>2003 – 2008</td>
<td>DOB, ZIP, SSN</td>
<td>Patient care and financial level data from CT-licensed outpatient surgical centers</td>
</tr>
<tr>
<td>Emergency Department (CHIME)</td>
<td>Flat file</td>
<td>14.5m records</td>
<td>1996 – present</td>
<td>DOB, ZIP, SSN</td>
<td>Patient care and financial level data from all members of CHIME</td>
</tr>
<tr>
<td>Cardiac Registry</td>
<td>SPSS</td>
<td>17k records</td>
<td>2004 – 2008</td>
<td>DOB, ZIP</td>
<td>Patient care level data from five acute care hospital cardiac programs</td>
</tr>
<tr>
<td>Hospital Reporting System</td>
<td>SQL</td>
<td>30 hospitals</td>
<td>1992 – present</td>
<td>Not Applicable</td>
<td>Not a patient level database</td>
</tr>
<tr>
<td>Registry of Exempt Facilities</td>
<td>Access</td>
<td>2k records</td>
<td>1994 – present</td>
<td>Not Applicable</td>
<td>Not a patient level database</td>
</tr>
<tr>
<td>Certificate of Need Database</td>
<td>SQL</td>
<td>Over 7k dockets</td>
<td>1973 to present</td>
<td>Not Applicable</td>
<td>Not a patient level database</td>
</tr>
</tbody>
</table>

*Targeted for initial inclusion in CHIN
University of Connecticut Health Center

The University of Connecticut Health Center (UCHC) maintains a large repository of healthcare data on Connecticut residents. Databases at UCHC include patient care information for those treated at John Dempsey Hospital as well as myriad research databases containing diverse information such as clinical data, behavioral surveys, and genomic information. The two databases that have been targeted initially for CHIN are the Neonatal Information System which holds all the data on infants admitted to the Neonatal Intensive Care Unit at JDH, and Quick Recovery, which is the UCHC Dental Clinic record system.

Quick Recovery, which runs on a proprietary platform, contains data from 2003 to the present on patient scheduling, diagnoses, treatment plans, dental procedures, financials and reporting for Connecticut residents seen at the UCHC dental clinics, Burgdorf/Bank of America Health Center, and other satellite clinics. Personal identifiers such as name, ZIP and Social Security Number are included in the database. 14

The Neonatal Information System (NIS), which also runs on a proprietary platform, contains data from 1995 to the present (with some additional data back to 1986) on infants admitted to the NICU at JDH. NIS serves as the clinical database for the NICU and is also used to fulfill mandatory reporting to DPH as well as benchmarking with other NICUs internationally and internally for quality assurance assessment. It includes information on patient and parent demographics, diagnoses, and clinical care received. NIS includes identifiers such as patient and parent name, dates of birth, SSN, and ZIP code, as well as patient medical record and account numbers. The current version of NIS used at UCHC is DOS based, however a SQL based version is available and may be utilized in the near future. 15

Additionally, the IDX system serves as JDH’s clinical database which holds patient demographics, admission, transfer and discharge records, diagnosis and treatment codes, as well as billing and insurance information for all patients seen at UCHC. IDX runs on DB2 and contains data on 1.3 million inpatients and outpatients seen at UCHC from 1994 to the present. Identifiers such as name, ZIP code and SSN are included in the database, although SSN is not found for all individuals. IDX also contains a unique patient identifier which is assigned to each individual at their first encounter in the JDH. 16

Table 5. Health and Human Services Data: UCHC

<table>
<thead>
<tr>
<th>Database</th>
<th>Platform</th>
<th>Size</th>
<th>Timeframe</th>
<th>Identifiers</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonatal Information Systems*</td>
<td>Proprietary</td>
<td>Over 10k Individuals</td>
<td>1995 - present</td>
<td>Patient and parent name, SSN, ZIP, DOB</td>
<td>Clinical database for UCHC NICU</td>
</tr>
<tr>
<td>Quick Recovery*</td>
<td>Proprietary</td>
<td>Over 200k individuals</td>
<td>2003 - present</td>
<td>Name, SSN, ZIP</td>
<td>Clinical database for UCHC Dental Clinics</td>
</tr>
<tr>
<td>IDX</td>
<td>DB2</td>
<td>1.3 m individuals</td>
<td>1994 - present</td>
<td>Name, ZIP, TOO number, SSN (not always present)</td>
<td>UCHC hospital database</td>
</tr>
</tbody>
</table>

*Targeted for initial inclusion in CHIN
III. A Solution to the Data Integration Needs of the State of Connecticut

The agency databases described above hold a wealth of data that currently serve many specific business needs. These data, however, remain highly fragmented, both within and between agencies, thwarting efforts at collaboration that crosses system or agency boundaries. For example, DCF clients who age out of the system may still benefit from services administered by other agencies such as DDS, DSS, and DMHAS. These agencies have business interests in data sharing for tracking and follow up. Currently these business needs and many others go unmet due in part to significant technical roadblocks, and the information sharing that does occur is labor-intensive and consumes considerable State resources.

The technical roadblocks to data sharing are primarily twofold. First, databases holding valuable information run on disparate platforms across, and even within, state agencies. For example, the 30 databases described in depth above run on more than a dozen different platforms. The lack of standardization and interoperability across systems makes obtaining and integrating data extracts from multiple sources difficult. Second, there is no common state identifier, nor for that matter are there standards for collecting demographic or identifying data. For example, while many databases include name or Social Security number as identifiers, some use identifiers that may be unique to that database or that are shared with only a select group of other databases. With different databases utilizing different identifiers, matching individuals across databases is extremely challenging.

The core mission of CHIN is to allow access to de-identified health information from across state agency databases for policy makers, researchers, and government officials. To solve the technical barriers identified above and to comply with all relevant statutes and regulations governing access to data without adding additional burdens to agency personnel, a fully functional, production network needs to have the following attributes and capabilities:

- Allow shared access to health-related data from heterogeneous data sources, providing tools for users to browse data elements and their definitions from these data sources and build queries using these data elements.
- Allow agencies to retain control over access to data.
- Have minimal impact on business practices.
- Enable creation of data sets and reports that integrate data at the level of the individual record by using cutting-edge approaches to record linkage.
- Ensure secure network communications
- De-identify data before providing to the end user.

A comprehensive study conducted by a University of Connecticut task force in 2005 concluded that the most feasible and efficient approach to achieve this objective is via a federated, metadata-driven network that securely integrates data records at the individual level. The federated approach, employed on the national level by CDC’s Public Health Information Network (PHIN), facilitates data sharing by defining a common set of standards & protocols and providing middleware software components to enable messaging and data exchange. Although other approaches to data integration – integrated systems, such as Missouri’s MOSAIC, or data warehouses, such as Illinois’s Chapin Hall – are capable of solving many of the problems identified above, they typically entail substantially higher costs and present more limited functionality to end users than do distributed systems. Further information on alternative approaches to data integration can be found in the CHIN Needs Analysis report which is available at http://publichealth.uconn.edu/CHIN.php.
Technical Framework for Implementation of CHIN

Because of its security, efficiency, scalability, and capabilities for safeguarding privacy and confidentiality, the CHIN platform proposed in this report utilizes the federated approach to data integration. This architecture allows users to access diverse data as if it were a single source by utilizing a distributed database architecture with record integration, query, and privacy capabilities in a secure networked environment. This technology framework provides authorized end users with the benefits of a data warehouse while avoiding much of the high maintenance overhead of a physical data warehouse.

As stated above, a key barrier to effective public health research using existing data is the lack of open, interoperable, transparent technology solutions that address bottlenecks in information integration and business process management. In striving to overcome these challenges and deliver solutions that improve research and evaluation capabilities while also assuring regulatory compliance, the proposed network relies on standards-based, open, service-oriented architectures, using federated grid architectures wherever possible. We believe this is the most sustainable, scalable long term approach that provides tremendous value to the State by reducing technical and operational barriers to participation.

Privacy, Security and Record Linkage for Secure Data Sharing in a Federated Environment

We have identified four critical implementation challenges for CHIN:

1. **Record Linkage** to enable linking of an individual’s records across data sources, which facilitates comparison and analysis for patient care and public health studies.
2. **Data De-Identification** to minimize sharing of protected health information (PHI); typically done for the databases at local sites in accordance with HIPAA and other federal and state privacy and confidentiality guidelines.
3. **Secure Communication** including providing tools for user authentication, ensuring message Integrity and maintaining user privacy.
4. **Access Control** to provide authorized users with selective access to specific datasets; this includes setting up user roles and privileges for specific studies, data sources and types of data they can access, query or update.

Each of these challenges is addressed by functionality and services residing on various nodes of the overall CHIN network. The overall infrastructure incorporates solutions for secure communication and access control. In addition to the technical controls built into CHIN, legal and regulatory restrictions, combined with appropriate administration and oversight, provide a highly capable security infrastructure. This multi-layered approach will minimize the risk of problems such as unauthorized access and the unintentional release of identifiable information. A more detailed discussion of each of these challenges is provided below.

**Record Linkage**

CHIN requires a means for linking individual records across data sources to allow for comparison in patient care or research studies. Record linkage is the process of finding records that refer to the same entity (in CHIN’s case, a person) in two or more independent data sources that do not have common unique identifiers. Individual records can be linked either by: 1) unique identifiers such as social security numbers or some widely shared medical, national, or universal identifiers, or 2) comparing ensembles of non-unique identifying data items such as name, sex, zip code and date of birth in combination. A range of identifier matching solutions are being developed for CHIN and will be used dynamically based on knowledge of the identifier and the nature of data queries themselves.
While record linkage requires the transfer and comparison of identifying fields from separate data sources that may contain personally identifiable information (PII), CHIN’s record linkage approach ensures that the system does not transmit PII beyond the minimum necessary for record linkage, and does not persistently store PII on any server outside of the agency that owns the source record. CHIN utilizes a Trusted Broker solution, which requires that identifying data are revealed only to a single component (the Trusted Broker). The Trusted Broker, a component of the CHIN Main Module, builds and maintains a secure index with anonymous, de-identified pointers (called record locators) to agency subject records using record linkage algorithms. The record linkage functionality requires transmission of PII from agency sources to the Trusted Broker. In this secure module temporary aggregation of PII data from records across multiple agencies is used to build and/or update an index of internal, non-identifiable record locators used by CHIN to reference matched and non-matched individuals across CHIN data sources. The linkage operation will generate record locator IDs based on exact matching and probabilistic matching using PII fields available in each data source. These locators are centrally stored in the Trusted Broker, with the PII data discarded once the linkage operation is complete; no PII data are persistently stored in the Trusted Broker or any other central location. The Trusted Broker also returns a corresponding local agency record locator for each subject record which is stored both centrally and within the CHIN agency module. Record locator IDs are not accessible to users of CHIN.

Data De-Identification

In CHIN, data fields that contain PII from participating agency databases are used only for record linkage and are not accessible as part of a user query. The architecture is designed to prevent information transmitted for record integration from being used to identify subjects. CHIN has a change management component that modifies operations when a change is executed to ensure that PII and other sensitive data are protected. For instance, if an Agency Supervisor removes a data source, or makes a data element no longer accessible, then existing queries that utilize the data element(s) in question are disabled.

Secure Communication

Data sources in CHIN share sensitive data and must be encrypted. CHIN utilizes high-grade 256-bit SSL encryption for all connectivity between the application and the user. In addition, each participating agency uses a secure connection to the Main Application that ensures the authentication and authorization of the site. The connection is done via a secure mechanism for site-to-site encryption, such as VPN or connection via certificate-based SSL connection to a proxy server.

Access Control

Four major sets of users are served by CHIN. End users -- researchers, agency personnel, and policymakers -- can query and retrieve data. Agency IT personnel manage the configuration, connectivity, & security of their data on the network. UCONN IT personnel support and administer centralized components of CHIN (i.e., record linkage engine, query interface, and system security architecture), and support agency IT personnel. Finally, IRB and administrative decision-makers oversee the usage of the network, set policies for privacy, confidentiality, network access, and use of data, and review system audit logs and usage reports.

-------------------

1 The PII referred to consists of data fields within CHIN explicitly labeled as PII data fields. Other fields (such as clinical records in narrative free text form) may contain unstructured PII that, if they are made available in CHIN, will need to be stripped out by other automated or manual means. These cases will be covered more comprehensively in a future version of CHIN.
CHIN relies on multiple layers of access control to manage users and ensure security and privacy:

- **Authentication** - Each user must log in to the application with a unique username and password. No user can access the system without a valid, approved user account. The password must have a minimum length of 8 characters and have at least 1 upper case character, 1 lower case character, and 1 number.

- **Role Based Access Control** - A user’s privileges are restricted according to his/her role. Each user is restricted to a specific role relative to the application and only accesses certain functionality and data defined by that role. Three roles are defined for the system:

  o **CHIN Host Site Administrator** – Manages and monitors the CHIN Main system as a whole, including log review, user management, query review/approval, and assisting agency personnel with data source management. User administration is an important part in the governance process to safeguard access to the application. The Host Site Administrator can access all user profiles through the user management screen of the user interface, and can create, modify, activate, deactivate and assign roles to user accounts.

  **CHIN Agency Administrator** – Manages the availability of a specific data source in CHIN and sets replication schedules for updating the in-memory cached database from the source data. Also manages the availability of data elements from a specific data source for a participating agency. The Agency Administrator decides what data elements can be included in the Metadata Registry and can be used in building queries.

- **Data Consumer** – Builds and executes queries for the purpose of data analysis and evaluation.

- **Data Source Control** - A key design concept of CHIN is distributed control – i.e., the data source of a participating agency should be under the control of the supervisor from that agency. CHIN allows an Agency Administrator to control the accessibility of his/her own data source. Similarly, the Agency Administrator can control access to individual data elements within a data source. An Agency Administrator may control the accessibility of the data elements within a data source by modifying whether the element is viewable in the metadata registry and whether it can be selected for inclusion in a query.

- **Query Execution** - A query created by a user must be reviewed and approved by the CHIN Host Site Administrator before it can be executed. CHIN requires strict guidelines regarding usage and privacy, which can be enforced as part of a query submission and review process.

**CHIN Functionality**

Users of CHIN are provided with services that enable them to Search, Access, Connect, Evaluate, and Administer within the CHIN Network.

**Search** – Authorized users may search the system in three ways:

1. Search a directory of data sources and data elements accessible in the system, with descriptions of available data fields and number of records in each data source.
2. Search a catalog of public health contacts for each agency and data source.
3. Visually design queries using a hierarchically structured library of data elements.

**Access** – Users may execute their query and download de-identified data (once authorized by an administrator) in MS Excel format or as tab-delimited flat files for import into statistical analysis programs.
Connect – Agency IT administrators may publish information about their data sources to the CHIN Registries in order to make them accessible to users. The publishing process includes establishing data source connectivity, defining data elements, configuring database mappings, setting access policies, and providing metadata on the source to the Registry.

Evaluate – The system will track usage patterns. It allows user annotations on data quality, use of standardized vocabularies, and query effectiveness. These data allow for more effective review and iterative improvement of the system.

Administer – Administrative personnel may manage user accounts, review and authorize queries, and audit system usage.

Architecture

The proposed network relies on the development principles of modular design and standards-based, federated, interoperable architectures, providing tools and fostering user communities that facilitate collaboration. CHIN consists of two modules. The first, the CHIN Main Module, is a server at the host site (UCHC) used for processing queries and brokering the match of records from multiple databases using a “trusted” or “honest” broker paradigm. The Main Module also provides query building, metadata browsing, and system management interfaces.

A second module, the CHIN Remote Module, is a server running at each agency site and is used to provide access to approved data sources. The CHIN Remote Module is managed via a user interface hosted on the Main Module. This allows agency users to manage the data sources and data elements that are provided to CHIN users. CHIN supports one Remote Module per data source included in the system.

CHIN has three key architectural attributes:

1. **Distributed** - data are collected from multiple remote data sources
2. **Hierarchical** - control of the system is hierarchical, the system runs checks on availability of the data sources and authorization of access prior to running any query. This process is controlled from the main module located at the central hosting site
3. **Dynamic** - data sources may change their definitions, or an administrator may remove the right to run a data query. The system will detect these changes and adjust accordingly.

As depicted in the high-level CHIN architecture diagram presented below, there are five key components of the CHIN architecture:

1. **Trusted Broker** – translates a subject remote unique identifier (GUID-based) into an application-level subject unique identifier. By having a subject uniquely identified at the remote location level, CHIN can associate a set of algorithms that will clean and standardize the record before it is sent to the main application. In this way CHIN can individualize how a subject record is processed remotely based on the specific data definition for a data source.
2. **Metadata Registry** – a central repository for all data elements and their definitions that are authorized for queries by agency managers.
3. **Query manager** – creates, retrieves, updates, and deletes queries.
4. **User manager** – creates, retrieves, updates, and deletes user profiles.
5. **Supervisor manager/Local supervisor** – handles both the operational and technical aspects of communicating to remote locations. The CHIN “supervisor” or central administrator acts as overseer and facilitator, while local agency administrators retain control over the availability of the data.

**CHIN ARCHITECTURE AND DATA FLOW**

**Access to Aggregate Data**

While access to raw data that have been integrated at the record level is essential for rigorous research and evaluation, key stakeholders, such as legislators, policymakers, local health directors, and community advocates, would benefit greatly from enhanced access to aggregate data for purposes of planning, monitoring and resource allocation. Such data would offer users the opportunity to obtain basic population statistics, such as means and percentage distributions, on critical health indicators and to assess differences by key demographic characteristics, such as gender, race/ethnicity, age, and geographic area (e.g., county, town, zip code).

Although not the focus of our current efforts to test the feasibility of a robust approach to data integration, the underlying CHIN architecture provides the platform necessary to enable web-based access to a rich array of indicators from multiple data systems across multiple state agencies. The necessary prerequisites for providing this functionality within CHIN are:

1. consensus on the type and form of aggregate data that should be presented;
2. development of an interface for accessing these data; and
3. acquisition of funding to support the development of this functionality within CHIN.
Community consensus

Regarding the first prerequisite, engagement of community stakeholders and agency personnel who would benefit from access to aggregate health data is an essential first step in developing such a plan. Ideally this process would include numerous participants, such as members of the Connecticut Association of Directors of Health, policymakers in the legislative and executive branches, and representatives from Connecticut nonprofit agencies and foundations, for whom access to such data would enhance planning, monitoring, and grant opportunities. These efforts should also include agency personnel who can assist in prioritizing the information available and address issues of data quality and other limitations. Moreover, this effort should capitalize on current efforts to provide access to aggregate data, such as the Department of Public Health’s Connecticut Health Information Reporting System (CHIERS). This web-based query system, which is currently in the pilot phase, uses extracts from HIP-Kids or other databases that are imported into CHIERS to provide summary community health statistics based on user-defined queries.

Development of a web-based query system

The CHIN architecture provides a number of options for developing this functionality. Existing open-source applications providing robust analytic tools and visualization functions, such as the UMass Lowell-led MICO-Viz Consortium for collaborative web-based visualization, could be incorporated into the CHIN platform. Alternatively, a custom application could be developed. The design and deployment of a custom application could potentially be expedited by adaptation of the aggregate reporting and analytic tool currently being developed for OpenClinica, a web-based clinical trials research platform created by our technology partner Akaza Research in collaboration with the University of Connecticut Health Center. Components of OpenClinica are currently being used in several high profile research collaboratives, including the National Repository for Autism Research (NDAR) and the Cancer Bioinformatics Grid (caBIG), and by the World Health Organization. OpenClinica uses many of the open-source components included in the CHIN application and could be readily adapted for use in this environment. Third, many commercial software packages offer this functionality (e.g., SAS), but would require ongoing licensing or maintenance fees and may offer limited potential for customization.

Funding

In addition to funding provided by the State of Connecticut, this functionality could be supported by federal initiatives promoting access to public health data. The Centers for Disease Control and Prevention has supported many state efforts to build public health capacity in this area through its Assessment Initiative. It is important to emphasize here that the open source nature of CHIN greatly enhances opportunities to access research and development funding from federal sources that have made the development of such systems a priority. Local foundations, particularly grant-making organizations that rely on access to such data to identify priorities and shape local and regional health policy, may provide financial support for efforts to improve data access and data quality.
Current Status of the CHIN Initiative
With funding from the General Assembly for FY 2008 and 2009, the Connecticut Health Information Network has moved from the Prototype phase, funded by the University in 2005 and 2006, to a Limited Production deployment that has allowed the University to conduct a formal assessment of the cost and feasibility of deploying a fully functional network. The current phase of the initiative is focused on evaluating, expanding, and refining the current Limited Production system to prepare for and assess the feasibility of a Full Production launch. Although some of this work remains ongoing, milestones achieved to this point or scheduled to be met by the end of the fiscal year include:

* The design, deployment, and testing of the main application at UCHC;
* The design, deployment, and testing of the remote application allowing CHIN to connect to data sources at UCHC and DOIT;
* Implementation of features to address scalability, security, and record linkage issues
* Testing of query infrastructure upgrade to handle data sources with 50,000+ records
* Addition of support for multi-table relational database mapping and flat file mapping
* Upgrade of Metadata Registry and Contributor Management capabilities to support automated registry management, publishing & updates by contributor administrators and more granular data security policies
* Enhancement of matching engine to conduct dynamic data source matching and utilize additional identifying fields
* Incorporation of system management and monitoring tools to ensure CHIN service reliability
* Establishment of live connections to 5 databases at the agencies and extracts from 2 other databases;
* Completion of formal testing to ensure connectivity, accuracy, performance, and usability.

Recommendation: We recommend production deployment of the CHIN infrastructure described above for the five agencies named in Public Act 07-02, in accordance with DOIT’s Systems Development Methodology and subject to negotiation of Memoranda of Agreement and, where applicable, Business Associate Agreements with the Center for Public Health and Health Policy at the University of Connecticut Health Center, which will serve as CHIN administrator. We also recommend expansion of this network to include other state agencies as funding and federal regulations permit.
IV. Inventory of Health-Related Data Aggregation and Data Access Improvement Efforts

The recognition of the need to improve data access and to overcome data fragmentation in the State is reflected in the numerous data access improvement and data aggregation efforts that are currently underway. These efforts are inventoried below. Although valuable in their own right, most of these efforts are not actual data integration initiatives -- i.e., they do not attempt to construct systems designed to automate the case-level integration of patient or client data. Virtually all focus primarily on increasing the accessibility of individual data systems, or on providing access to aggregated data for reporting and monitoring. Most are limited in scope, focusing exclusively on the improvement or integration of a specific set of databases, and are not developing technologies that will enable platform-independent integration across a range of contributors. Thus, we view these efforts as an important part of an overall data quality and data access improvement strategy that would have as its centerpiece the connectivity and functionality that is envisioned in CHIN.

Single agency efforts

Department of Children and Families

The Department of Children and Families currently has several data initiatives underway that improve access to or the quality of their data. Additionally, they have two data integration efforts underway. All of these data initiatives are aimed at improving the utility and functionality of data held in the LINK systems. For example, the Behavioral Health Tracking System has implemented an enhancement to provide daily tracking of mental health admissions of DCF clients. This enhancement uses data provided by Value Options, a DCF contractor, which is a combination of daily census data from LINK with reports on inpatient mental health recipients. DCF also has a data mart in which numerous data elements of LINK are aggregated in order to improve child welfare reporting to DCF social workers and management of children in placement, allegations of abuse and neglect and investigation status.

Two data initiatives allow DCF to share data with other organizations. DCF provides aggregate data from LINK to Chapin Hall, which provides aggregate child welfare data regionally and nationally. DCF also has the IV E project which is a data quality improvement effort aimed at maximizing financial recoveries claimed under Section IV E of the Social Security act. In this project data from LINK interfaces with judicial data sources. Lastly, the Results Oriented Management (ROM) project is run off of the Reports_CT database (which obtains data from LINK). ROM has been implemented as a third-party system which tracks child welfare measures that have been mandated by the Juan F Decree. These measures include workload, visitation and outcome aspects of child welfare. ROM aggregates data and makes it available to social workers statewide.

On hold is DCF’s MedLink Project, a data initiative focusing on improving DCF’s capacity to report on the health status of youth in their care. This project will couple a more user-friendly Medical Profile with increased reporting capacity to ensure that the health care needs of DCF clients are being met. Additionally, this project would allow the aggregation and reporting of related variables. Data for the MedLink project also originates from LINK. 18
**Table 6. DCF Data Access Improvement /Aggregation Efforts**

<table>
<thead>
<tr>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication Module to Link (MedLink)</td>
<td>On Hold</td>
<td>Increases the capacity to report on how children fare medically or to compare their care to established national health benchmarks in order to ensure that fundamental health needs are being met; provides more efficient packaging of children's health data so that it can be easily understood and communicated by agency personnel; allows the aggregation and reporting of key variables.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Behavioral Health Tracking System</td>
<td>Implemented</td>
<td>Daily tracking of mental health admissions by type showing length of stay maintained by the ASO.</td>
<td>Q</td>
</tr>
<tr>
<td>Data Mart</td>
<td>Implemented</td>
<td>Numerous aggregations of key elements of LINK to improve reporting. Provides a single source of information which is updated daily and used to generate various child welfare reports. Provides reports to child welfare social workers and DCF management concerning children in placement, allegations of abuse and neglect, and investigation status.</td>
<td>A</td>
</tr>
<tr>
<td>Chapin Hall</td>
<td>Implemented</td>
<td>Data are aggregated without HIPPA identifiers to allow sharing of child welfare information regionally and nationally.</td>
<td>A</td>
</tr>
<tr>
<td>IV E</td>
<td>Development</td>
<td>Improve data quality using DSS and Judicial interfaces. The initiative is to maximize financial recoveries claimed under Section IV E of the Social Security act.</td>
<td>A, Q, I</td>
</tr>
</tbody>
</table>

A= Data Access Project; Q= Data Quality Project; I= Data Integration Project

**Department of Developmental Services**

The Department of Developmental Services currently has no data integration efforts underway, but has been engaged in two data efforts aimed at improving the access to or quality of their data. DDS has recently implemented a platform migration of eCAMRIS in order to improve its data storage and access. Additionally, web-based applications are being developed for Level of Need in order to improve its utility.19

**Table 7. DDS Data Access Improvement /Aggregation Efforts**

<table>
<thead>
<tr>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Need</td>
<td>Production</td>
<td>Web-based application that allows client Level of Need assessment data online.</td>
<td>A</td>
</tr>
<tr>
<td>eCAMRIS Migration from Unisys to .NET SQL</td>
<td>Implemented</td>
<td>Improvements to management of services for DDS consumers. Provides a more robust and versatile centralized data repository with a structure that supports seamless data access from various rapid application development (RAD) products.</td>
<td>A, Q</td>
</tr>
</tbody>
</table>

A= Data Access Project; Q= Data Quality Project
Department of Public Health

The Department of Public Health has several data access improvements, data quality improvements and data integration efforts that are currently underway or completed. Examples of these efforts include the EMS/Trauma Registry System, the PHIN Messaging (PHINMS), the Health Alert Network, and the CT Electronic Disease Surveillance System (CT EDSS) and CT Environmental Public Health Tracking Network. All are aimed at improving the capture and transport of sensitive health information collected and reported on by DPH.

DPH data integration efforts include the creation of a statewide health information technology plan which is currently in the business requirements phase of development. The plan includes the development and construction of a Health Information Exchange for the State.

Under construction is the HIP-Kids data warehouse that links databases within DPH to create a single child health profile. The expanded data warehouse will add to the existing core CHIP databases (birth records, newborn hearing screening birth defects registry, CYSHCN/Medical Home Initiative,) by adding information from the immunization registry, cancer registry, death records, lead surveillance, WIC, CEDSS.

Already implemented are three sets of linkages that are done annually to connect DPH databases. The infant birth and death linkage allows for the monitoring of trends of infant and maternal risk factors related to infant mortality. This project is also aimed at improving race/ethnicity classifications for infant mortality data. The linkage of maternal deaths with birth and fetal death records allows for the identification of pregnancy-related and -associated deaths. Lastly, the birth records and hospitalization linkage allows calculation of hospital delivery costs for newborns and calculates cost-associated missed opportunities to prevent adverse outcomes.

Also already completed is DPH’s pilot project, the Connecticut Health Information Reporting System (CHIERS). CHIERS is a web-based data query system into which extracts from HIP-Kids or other databases can be imported. CHIERS can then be used to provide summary community health statistics based on user-defined queries. The pilot implementation of CHIERS includes town-specific data including: population counts by age-sex-race-ethnicity; birth risk factors and outcomes; childhood blood lead and newborn hearing screening.

On hold is DPH’s Immunization Registry project which aims to transfer the system to a web-enabled platform. Additionally, this project aims to both expand registry access to more providers and to allow for data exchange with other systems.20

Table 8. DPH Data Access Improvement /Aggregation Efforts

<table>
<thead>
<tr>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Information Technology Plan</td>
<td>Business Requirements, Developing RFP.</td>
<td>Statewide health information technology plan to include the development and growth of Health Information Exchanges (HIE) in Connecticut.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Laboratory Information Management System (LIMS)</td>
<td>On Hold</td>
<td>Provides real-time laboratory results reporting to private sector, state and federal health officials to prevent illness, disease and death.</td>
<td>A, Q</td>
</tr>
<tr>
<td>EMS/Trauma Registry System</td>
<td>Implementation</td>
<td>Capture of statewide data from all 911 calls and collection of trauma data from all hospitals with trauma service.</td>
<td>A</td>
</tr>
<tr>
<td>PHIN Messaging (PHINMS)</td>
<td>Construction</td>
<td>Secure data transport system from the CDC. This project has 3 subcomponents: the PHIN Messaging System (PHIN MS) allows local, state, and federal public health professionals and organizations to rapidly and securely transmit sensitive health information over the Internet to other local, state, and federal public health professionals and organizations.</td>
<td>A</td>
</tr>
<tr>
<td>Project</td>
<td>Status</td>
<td>Description</td>
<td>Classification*</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>HIP-Kids (Designed on the expansion of the Child Health Informatics Profile (CHIP))</td>
<td>Construction</td>
<td>Data warehouse that links databases within DPH to create a single child health profile. The expanded data warehouse will add to the existing core CHP databases (birth records, newborn hearing screening birth defects registry, CYSHCN/Medical Home Initiative,) by adding information from the immunization registry, cancer registry, death records, lead surveillance, WIC, CEDSS.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Connecticut Health Information Reporting System (CHIERS)</td>
<td>Pilot Project Completed</td>
<td>This web-based data query system (WDQS) is modeled after the Missouri MICA system. Data extracts from HIP-Kids or other databases can be imported into CHIERS and used to provide summary community health statistics based on user-defined queries. The pilot implementation of CHIERS includes town-specific data including: population counts by age-sex-race-ethnicity; birth risk factors and outcomes; childhood blood lead and newborn hearing screening.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Infant birth and death records</td>
<td>Implemented</td>
<td>Annual linkage that provides data for the fetal and infant mortality survey. Monitors trends of infant and maternal risk factors related to infant mortality, and improves race/ethnicity classifications for infant mortality data.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Maternal Deaths</td>
<td>Implemented</td>
<td>Annual linkage of female deaths with birth and fetal death records to identified pregnancy-related and associated deaths.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Birth records and Hospitalizations data</td>
<td>Implemented</td>
<td>Allows calculation of hospital delivery costs for newborns; calculates cost-associated missed opportunities to prevent adverse outcomes.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Electronic Vital Records Solution</td>
<td>Phase 1: Implementation</td>
<td>Automation of the reporting, collecting and processing of death, birth, marriage and civil union certificates. Conversion of all certificates issued since 1935 is being considered.</td>
<td>A</td>
</tr>
<tr>
<td>CT Immunization Registry and Tracking System (CIRTS)</td>
<td>On Hold</td>
<td>Transition of registry from FoxPro to a web-enabled system to expand registry access to more providers and allow for data exchange with other systems such as Medicaid.</td>
<td>A,Q, I</td>
</tr>
<tr>
<td>CT Electronic Disease Surveillance System (CT EDSS) and CT Environmental Public Health Tracking Network</td>
<td>Main products: Production</td>
<td>Web-enabled transactional database system. Maven system receives data via PHIN maven and parses it into correct data fields. Data are received from laboratory and provider reporting as required by state health code. Areas of reporting include occupational health and condition tracking.</td>
<td>A, Q</td>
</tr>
<tr>
<td>Health Alert Network (HAN)</td>
<td>Design</td>
<td>Provides public health preparedness, response, and service on a 24/7 basis to ensure rapid and timely response.</td>
<td>A, Q</td>
</tr>
</tbody>
</table>
### Office of Health Care Access

The Office of Health Care Access currently has no data integration efforts underway, but has a data access and quality improvement project that was completed in 2008. OHCA’s Hospital Reporting System has been converted to incorporate new functionality and to facilitate data download and reporting capacities.

#### Table 9. OHCA Data Access Improvement/Aggregation Efforts

<table>
<thead>
<tr>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Reporting System</td>
<td>Implemented</td>
<td>Conversion of the Hospital Reporting System to incorporate new formulas and to improve downloads and reports.</td>
<td>A, Q</td>
</tr>
</tbody>
</table>

A= Data Access Project; Q= Data Quality Project

### Other State Agencies

Numerous other state agencies are also engaged in data access, data quality and data integration efforts. Most of these efforts are aimed at improving case management. DMHAS is involved in upgrading their Behavioral Health Information System (BHIS) to a more user-friendly system which also promotes HIPAA compliance. The upgraded system is intended to better serve DMHAS clients through improved functions that increase data quality and tools to providers. Additionally, the Recovery Management System will interface with BHIS to import client and service information and provides for direct electronic entry of treatment plans and progress notes.

DSS is also involved in two projects that are a part of a Health Information Exchange in DSS: the Client Pharmacy Reimbursement and Medical Records System and the Electronic Health Records Project (oh hold). The Client Pharmacy Reimbursement and Medical Records System will create medical records for DSS clients that show the historical use of medications. The Electronic Health Records Project is intended to function as a primary care case management program. This project adds clinical information into existing systems that collect claims information and also creates mechanisms for physicians to share that clinical information.
<table>
<thead>
<tr>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Project</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMHAS</td>
<td>Behavioral Health Information System (BHIS) Upgrade</td>
<td>Business Requirements Upgrade existing system to Avatar. Promotes HIPAA compliance, the assurance of legible and consistent medical records information, the availability of tools to assist clinical providers, the capture of client, clinical and service level data to trigger billing and support client activity through DMHAS and the need to have a user-friendly system to minimize the time of clinical providers in recording required information and maximize their time for support of DMHAS clients.</td>
<td></td>
<td>A, Q</td>
</tr>
<tr>
<td>DMHAS</td>
<td>Web-based Mental Health Referral System</td>
<td>Production Web-based mental health resource referral for public. This site will provide directory information about assistance, programs, services, providers and legislation. Links to advocacy organizations and other important information.</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>DMHAS</td>
<td>Recovery Management System</td>
<td>Construction Electronic treatment plan linked to service delivery and progress notes. Interfaces with BHIS to pull in client and service information and provides for direct electronic entry of treatment plans and progress notes.</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>DSS</td>
<td>Electronic Health Records (including e-Prescribing)</td>
<td>On Hold Project adds clinical information to systems that currently collect claims information. Involves creation of electronic health records and exchange mechanisms to share clinical information amongst physicians.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>DSS</td>
<td>Client Pharmacy Reimbursement and Medical Records System</td>
<td></td>
<td>Project will create medical records for clients that show historical use of medications. Intent is to improve health care delivery, reduce prescription errors, provide reminders for refills, and improve drug formulations for patients. Project may also create beginnings of a Health Information Exchange (HIE) that could be useful to other agencies.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>DSS</td>
<td>Eligibility Management System</td>
<td></td>
<td>Welfare eligibility system. Includes 68 databases covering 1529 programs, millions of client records.</td>
<td>I</td>
</tr>
<tr>
<td>DVA</td>
<td>Patient Care and Patient Billing Systems</td>
<td>Business Requirements Replaces legacy systems with a HIPAA-compliant solution that modernizes the graphical user interface. Systems are used for admissions, doctor orders, appointments, pharmacy orders and invoices, financial reporting, and tracking Medicare and Medicaid billing.</td>
<td></td>
<td>A</td>
</tr>
</tbody>
</table>

A= Data Access Project; Q= Data Quality Project; I= Data Integration Project
Multi-agency efforts

There are currently 12 multi-agency data integration efforts active in the State of Connecticut other than CHIN. Some of these efforts focus on improved case management while others have a research and public policy focus. Four of the larger state projects include the Data Interoperability Initiative; Ready by 5, Fine by 9: CT’s Early Childhood Investment Initiative; Connecticut Data Quality and Access Initiative; and the Health Information Exchange.

The Data Interoperability Initiative is led by the Office of Workforce Competitiveness (OWC) and exists as a partnership between OWC, the Department of Mental Health and Addiction Services (DMHAS), and the Bureau of Rehabilitation Services (BRS). The initiative grew out of the Mental Health Systems Transformation Grant, the multi-agency working group on youth that resulted from the NGA Policy Academy: Transitioning Youth with Disabilities, and the Medicaid Infrastructure Grant. Agencies working on those projects recognized that challenges exist for data sharing between state human service agencies. The initiative focuses on improving data sharing between state agencies for both case management and analytical research.

The Connecticut Data Quality and Access Initiative is a public-private partnership that will offer through a web site a wide range of data sets to support state, regional, and local policy and planning efforts, beginning in the areas of early childhood, health, and economic development. The Initiative is a partner in a national consortium of community indicator projects operated out of the University of Massachusetts Lowell that is leveraging development funds to build innovative data visualization tools. This suite of tools, tailored to a range of user levels from novice to expert, will include linked mapping and charting with a high degree of rapid interactivity. The project is open to partnering with other sectors seeking an efficient platform for offering data at various geographic levels on the web.

Table 11. Multi-Agency Data Access Improvement/Data Integration Efforts

<table>
<thead>
<tr>
<th>Agencies</th>
<th>Project</th>
<th>Status</th>
<th>Description</th>
<th>Classification*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCF and CSSD</td>
<td>Juvenile Justice Strategic Planning Initiative</td>
<td>Planning</td>
<td>Electronic sharing of case specific information.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>DMHAS, DDS, 13 State Departments, Judicial Branch</td>
<td>Mental Health Systems Transformation Grant (SAMHSA Grant)</td>
<td>Planning</td>
<td>Develop electronic records and integrated systems to support cross agency care.</td>
<td>Q, A, I</td>
</tr>
<tr>
<td>DOC, DCF, DMHAS, Judicial Branch, OPM, Office of the Child Advocate</td>
<td>Multi agency working group on youth (MAWGY)</td>
<td>Implemented</td>
<td>Improve data collection, analysis and information sharing across agencies involved with incarcerated youth.</td>
<td>Q, A, I</td>
</tr>
<tr>
<td>DPH, DSS</td>
<td>Birth-Medicaid-WIC linkage</td>
<td>Implemented</td>
<td>Manual linkage of data annually to evaluate associations of infant and maternal risk factors and pregnancy outcomes.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>DSS, BRS, DDS, DMHAS, SDE, DOL, DCF, DOC, CSSD, Board of Ed Services for the Blind, Capital Workforce Partners (WIB)</td>
<td>NGA Policy Academy: Transitioning Youth with Disabilities</td>
<td>Implemented</td>
<td>Improved data interoperability related to services for transitioning youth with disabilities to independent adulthood. Clients provide information once across agencies, allows for cross agency comparisons, improves case management, allows for measurement of impact of state services, and improves coordination for provider payments.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Agencies</td>
<td>Project</td>
<td>Status</td>
<td>Description</td>
<td>Classification*</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>DSS Bureau of Rehabilitation Services, DDS, DMHAS, Board of Education and services for the Blind</td>
<td>Medicaid Infrastructure Grant (MIG)</td>
<td></td>
<td>Consolidate 9 databases to improve tracking the progress of individuals with disabilities in obtaining work or post secondary education. Databases include the Eligibility Management Information system (DSS), the Medicaid management information system (DSS), and the CT Automated Mental Retardation Information System.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Governor's Early Childhood research and Policy Council: K-12 Leadership, OPM, Mayors, CDA, OWC, CHEFA, CT DHE</td>
<td>Ready by 5, Fine by 9: CT's Early Childhood Investment Initiative</td>
<td>Planning</td>
<td>Improve health and development of young children; improve readiness for kindergarten and achieve early elementary school reading mastery. Linkage of databases, creation and consolidation of databases, interoperability standards.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>SDE, CT Department of Higher Education</td>
<td>PreK-16 Council</td>
<td></td>
<td>Statewide longitudinal Pre-K-16 data system.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>OWC</td>
<td>Mental Health Transformation (MHT) Data Interoperability</td>
<td>Problem Definition</td>
<td>Data sharing effort between DMHAS, DSS, BRS, DDS, the Governor’s Early Childhood Cabinet, and the United Way. Grew from MIG, MAWGY, NGA Policy Academy, SAMHSA initiatives.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>Early Childhood Education Cabinet (drawing data from SDE, DPH, DCF, DSS, DDS)</td>
<td>Connecticut Data Quality and Access Initiative</td>
<td>Development</td>
<td>Development of web site with state-of-the-art data visualization tools to facilitate dissemination and analysis of data on young children, early childhood services, health, and economic development.</td>
<td>A, Q</td>
</tr>
<tr>
<td>DSS</td>
<td>Data Warehouse</td>
<td>Phase I: Production Phase II: Business Requirements.</td>
<td>Allows access to program data with advanced analytical capabilities in order to improve revenues, constrain costs, improve ability to monitor program effectiveness, decentralize program reporting and collections of related data on providers and clients. Data will be downloaded every two weeks and shared by DSS, DCF, AG, Chief State’s Attorney, OPM and Leg. OFA.</td>
<td>A, Q, I</td>
</tr>
<tr>
<td>DCF, DSS</td>
<td>Connecticut Behavioral Health Partnership</td>
<td>Implemented</td>
<td>Under third party management. Uses data from DCF and DSS to improve the access to and quality of behavioral health services for eligible DCF and DSS clients.</td>
<td>A, I</td>
</tr>
</tbody>
</table>

A= Data Access Project; Q= Data Quality Project; I= Data Integration Project
Solutions offered by CHIN

CHIN provides a solution that the other data aggregation efforts cannot offer by providing a secure framework for the sharing and dissemination of de-identified data across health and human services agencies across the state. As a federated network of multiple databases that securely integrates records at the individual level, CHIN provides functionality that currently does not exist, and offer this in a cost-effective and non-disruptive manner. The CHIN architecture is highly scalable, having been specifically designed to easily, efficiently and securely grow to incorporate additional data sources as deemed appropriate by agency leadership and as allowable by law. Furthermore, CHIN is a software solution that does not limit participation to any specific platform(s), but can readily incorporate databases on disparate platforms (including “legacy” databases). CHIN also utilizes cutting edge integration algorithms to allow for the matching of records from databases that do not follow consistent standards for identifiers and other demographic fields. With an easy-to-use front end, CHIN provides a comprehensive and robust solution to facilitate research, evaluation, and program monitoring by a variety of user groups, and the system's capacity to support an interface through which aggregate data can be readily accessed allows it to provide utility to non-statistically oriented users.

Recommendation: Other multi-agency data integration efforts should be directed to collaborate with CHIN in developing their initiatives, and where appropriate, to capitalize on the state’s investment in the technologies developed for CHIN.

V. Privacy and Confidentiality

In order to explore the legal restrictions and confidentiality safeguards applicable to the personally identifiable information contained in databases to be included in CHIN, a CHIN Policy Workgroup was formed. The workgroup was made up of representatives from the agencies identified in the 2007 legislation, including the UCHC Center on Public Health and Health Policy, the Departments of Public Health, Developmental Services, and Children and Families and the Office of Health Care Access. A list of workgroup members is attached to this report as Appendix B.

Although CHIN will only provide de-identified data in response to approved queries, execution of queries will require participating agencies to provide identified data to CHIN for the purposes of record matching. Therefore, the workgroup explored the federal and state privacy and confidentiality statutes and regulations applicable to the various databases under consideration for inclusion in CHIN. It also reviewed the current policies and procedures of each agency regarding requests for identifiable information from its databases, including research proposals. The UCHC and the Departments of Public Health, Developmental Services and Children and Families each require all requests for access to data for research purposes to be submitted to the respective agency’s Institutional Review Board (IRB). (The Office of Health Care Access is prohibited by state statute from disclosing individually identifiable data for any purpose, and therefore does not have an IRB.) The workgroup also reviewed the policies and procedures of each IRB. A list of the federal and state statutes and regulations, and the IRB policies and procedures that were reviewed, are attached as Appendix C.

This section of the report reviews these state and federal privacy and security laws and regulations, as well as the institutional policies and procedures, relevant to individually identifiable data maintained by each agency included in the 2007 legislation. It provides the workgroup’s recommendations for addressing these provisions so as to promote the accessibility of integrated cross-agency data for program evaluation, planning, policy-making and research in a manner that respects the privacy and confidentiality of personal health information and complies with all relevant state and federal law.

Prior to reviewing these regulations it is important to emphasize the two critical conceptual pillars on which CHIN has been developed. First, all data released through CHIN will be de-identified in a manner that complies
with state and federal laws governing personally identifiable information. De-identified data requires a much lower threshold for release under most state and federal statutes. For example, de-identified data from education records subject to FERPA and from health records subject to HIPAA can be released without restriction. Second, the CHIN administrative core will not be engaged in any direct research activities. The administrative core will essentially serve as an information technology contractor providing services to the contributing agencies for the purpose of data integration and de-identification, under fully executed Business Associate agreements that are compliant with HIPAA. (All the agencies named in the CHIN legislation currently allow contractors providing IT services to have access to identified data). The exclusive provision of de-identified data for research, coupled with the separation of research and IT functions, enables CHIN to comply with some of the more stringent legal barriers to data sharing, e.g., FERPA restrictions articulated by the US Department of Education.

**Federal Laws**

**Health Insurance Portability and Accountability Act (HIPAA), 42 U.S.C. 201**

**HIPAA Privacy Rule, 45 C.F.R. parts 160 and 164**

The HIPAA Privacy Rule allows a covered entity to use or disclose health information that is de-identified without restriction and allows for the creation of databases for research purposes. The sole requirement is that the covered entity must first determine that the information has been de-identified using either statistical verification of de-identification or by removing certain pieces of information from each record covering eighteen elements as specified in Rule 45 C. F. R. §164.514(b). A HIPAA Covered Entity is permitted to engage other covered entities as Business Associates to de-identify PHI and to verify de-identification.

UCHC, DDS and DCF are Covered Entities under HIPAA, and their individually identifiable health data are considered protected health information subject to the privacy and security safeguards mandated in HIPAA. These include the obligation to keep such data confidential and secure, and to maintain a record of disclosures of such data. Provisions in HIPAA for sharing such data with Business Associates and researchers require all recipients of PHI to comply with the same HIPAA restrictions as the agencies themselves, and researchers may not re-disclose protected health information in any form that either reveals the identity of the individual to whom the information pertains or makes it possible to re-identify such individual.

DPH is a hybrid entity with its lab being the only component that is subject to the HIPAA restrictions. The rest of the DPH is a public health agency, as is OHCA, and is not subject to HIPAA.

**Protections for Human Subjects 45 C.F.R. Part 46**

UCHC, DPH, DDS and DCF require all requests for individually identifiable data for research to be submitted to the respective agency’s Institutional Review Board (IRB) in compliance with 45 C.F.R. Part 46, subpart A. Under these regulations, the Commissioner of each agency and the Executive Vice President of the UCHC must file assurances with the appropriate federal agency that the agency or the UCHC will comply with the federal requirements for the protection of human research subjects. Each IRB must certify that research proposals approved by it meet these requirements.

Although research proposals utilizing CHIN may be deemed to be exempt from IRB review because only de-identified information will be made available to the researcher (see 45 CFR section 46.101(b)(4)), IRB regulations require this determination to be made by the IRB of each agency.

**Recommendation:** Each agency should issue guidelines for human subjects’ review of data accessed through CHIN. In addition, the CHIN collaborative should seek guidance from the Human Subjects Protection Office of the Department of Health and Human Services regarding the protocol for IRB review of data accessed through
CHIN. Efforts to streamline the review process should be explored, perhaps including the creation of uniform IRB applications and a central IRB.


Data maintained by the "Birth to Three" system in DDS are educational records and are subject to restrictions under the Family Education Rights and Privacy Act (FERPA), 20 U.S.C. 1232g et seq., 34 C.F.R. part 99. This federal law prohibits the disclosure of personally identifiable information from an education record without the consent of the parent or student unless the disclosure meets specified exceptions set forth in 34 CFR Part 99.31. These regulations allow the disclosure of such information to organizations, including other government agencies, only for the purposes of developing, validating or administering predictive tests, administering student aid programs or improving instruction. Any such study must be conducted in a manner that does not permit the identification of parents or students by other persons, and the data must be destroyed when no longer needed for the study.

FERPA allows for the disclosure of de-identified data for research purposes without restriction. Recent rulings by the Family Policy Compliance Office of the US Department of Education indicate that disclosure of identifiable education records to other entities interested in matching records for research is not allowable (Letter to California Dept. of Education 2/18/2004; Letter to Pennsylvania Dept. of Education 2/25/2004; Memorandum from Dep. Secretary William Hansen 1/30/2003). However, these same rulings indicate that a contractor may conduct data matching as an “authorized representative” if under the direct control of the state educational authority.

*Medicare/Medicaid Intermediate Care Facilities for the Mentally Retarded (ICF/MR), 42 CFR part 483.410*

As a condition of participation in the Medicare and Medicaid programs for nursing facilities, DDS must comply with the conditions set forth at 42 C.F.R. part 483.410. These regulations require ICF/MR facilities, e.g., Southbury Training School, to keep all information contained in clients’ records confidential, and to develop policies and procedures governing the release of such information. Similar requirements pertain to the records of DDS clients receiving services under the Medicaid Home and Community Services Waiver.

**State Laws**

*Laws of General Applicability*

CGSA sec. 4-193 requires all state agencies that maintain personal data on individuals to protect such data from fire, theft, flood, natural disaster or other physical threat, and to inform their employees of this statute and any other state or federal statutes concerning the maintenance or disclosure of personal data by that agency. Sec. 4-196 requires the agencies to adopt regulations which conform to uniform standards adopted by the Attorney General’s office.

**Agency-Specific Laws, Regulations and Policies**

*Department of Children and Families*

Personally identifiable information maintained by the Department of Children and Families (DCF) is governed by CGSA section 17a-28. This statute makes all such information confidential and allows disclosure of such data only with the consent of the person named in the record, with certain specified exceptions. The exceptions listed in 17a-28 deal primarily with disclosures pursuant to reporting or investigation of abuse or neglect of a child, or to court proceedings involving such children. Unauthorized disclosure of confidential information from DCF files is punishable by fine and/or imprisonment.
Under 17a-28, information may be disclosed by DCF for research purposes only with written authorization from each person identified in the record and with the written authorization of the Commissioner of DCF or designee. Section 17a-101k makes the confidential information in the Central Registry of victims of child abuse or neglect subject to federal privacy and confidentiality laws and regulations.

**Department of Developmental Services**

The right of DDS clients’ to have their personal records treated as confidential has been established under CGSA section 17a-238, which requires that each person placed or treated under the direction of the Commissioner of DDS receive humane and dignified treatment with full respect for such person’s personal dignity and right to privacy consistent with such person’s treatment plan. Client records have been held to be “medical or similar files” under the Freedom of Information Act, the disclosure of which would constitute an invasion of personal privacy (Edelman, DKT# FIC 1997-089, Final Decision). These records are also protected by DDS regulation 19-570-5 and DDS provider licensing regulations which require licensees of the Department to maintain the confidentiality of client records (17a-227-1 et seq.). Probate records pertaining to guardianship or involuntary placement of individuals with mental retardation are required to be sealed under CGSA sections 17a-274 and 45a-668 et seq. These records may only be obtained through court order. In addition, parts of DDS client records may be subject to statutory “privilege,” such as doctor/patient, psychiatrist/patient or psychologist/patient (e.g., sections 52-146c et seq.)

**Department of Public Health**

In CGSA sec. 19a-2a(10), the Commissioner of Public Health is given, among other powers and duties, “...the power and the duty to .... Specify uniform methods of keeping statistical information by public and private agencies, organizations and individuals, including a client identifier system, and collect and make available relevant statistical information including the number of persons treated, frequency of admission and readmission, and frequency and duration of treatment.” The client identifier system is made subject to the confidentiality requirements of CGSA sec. 17a-688, which provides:

“(e) The commissioner may use or make available to authorized persons information from patients’ records for purposes of conducting scientific research, management audits, financial audits or program evaluation, provided such information shall not be utilized in a manner that discloses a patient’s name or other identifying information.”

In addition, many other DPH statutes contain their own references to the confidentiality of data collected under them. These restrictions vary depending on the individual statute and database. For example, CGSA 19a-7 designates DPH as the lead agency for public health planning in Connecticut and requires the Department to develop a multiyear state health plan. In connection with the development of this plan, the department is entitled to access hospital discharge data, emergency room and ambulatory surgery encounter data, data on home health care agency client encounters and services, data from community health centers on client encounters and services and all Office of Health Care Access data. The Commissioner is required to adopt regulations to assure the confidentiality of personal data and patient-identifiable data collected or compiled pursuant to section 19a-7.

Several sections in Title 19a require personally identifiable information to be kept confidential in accordance with section 19a-25. These include sections 19a-6e (Traumatic Brain Injury Registry); 19a-7h (Childhood Immunization Registry); and 19a-56b (birth defects information). Section 19a-25 deals generally with the confidentiality of records procured by the DPH and local health directors for the purpose of reducing morbidity or mortality. Such data “shall be confidential and shall be used solely for the purposes of medical or scientific research and....disease prevention and control.” Section 19a-25 also provides that “notwithstanding the provisions of chapter 55 (which deals generally with the obligations of all agencies with regard to personal data),
the DPH may exchange personal data for the purpose of medical or scientific research, with any other governmental agency or private research organization; provided such state, governmental agency or private research organization shall not further disclose such personal data.”

DPH has promulgated regulations 19a-25-1 through 19a-25-4 in implementation of section 19a-25. Regulation 19a-25-3 specifies the circumstances under which identifiable health data can be disclosed. Such disclosure must be limited to the minimal amount necessary to accomplish the public health or approved research purpose.

In addition to section 19a-25, separate statutes apply to the databases of vital records and statistics kept by the DPH, such as birth and death records. These records are covered by sections 19a-41 through 19a-45a.

Data acquired by DPH from outside sources may also be restricted by the source entity through language contained in a contract or Memorandum of Understanding (MOU). An example of this is the Vital Records Inter-Jurisdictional Agreement, which provides for the sharing of vital records information among the States, U.S. territories and Canadian provinces. Under this agreement, records of births and deaths can be sent by the jurisdiction where they occurred to the jurisdiction of usual residence. Death records can be sent to the jurisdiction of birth of the decedent. Receiving jurisdictions are required to abide by the restrictions of the sending district, which can vary by jurisdiction. Pursuant to this Inter-Jurisdictional Agreement, DPH considers death data received from other states to require approval from the sending state prior to release for research purposes.

The DPH lab may only disclose testing results to the provider who requested the testing and cannot disclose identifiable data for research purposes pursuant to sections 492.1109(e) and 493.2 of the Clinical Laboratory Improvements Amendments (CLIA) of the CDC regulations, and section 19a-36-D32(a) of the Regulations of Connecticut State Agencies. Therefore, all DPH data that is accessed for research purposes must be obtained from programs within the DPH, other than the lab, as permitted by the statutes and regulations governing those programs and their data.

Office of Health Care Access
OHCA is authorized under CGSA section 19a-613 to collect patient-level outpatient data from health care facilities or institutions, and under section 19a-654 to collect inpatient data from the state’s hospitals. Section 19a-654 requires that individually identifiable patient and physician data shall be kept confidential and only aggregate reports which do not identify patients or physicians shall be made available to the public. As a result, OHCA does not make personally identifiable data available for any purpose.

University of Connecticut Health Center
The University of Connecticut maintains numerous databases with individually identifiable patient health data, both in its role as a provider of medical, dental and hospital care, and in connection with its various bio-medical, medical and dental research projects. It is therefore subject to the privacy and security provisions of HIPAA with regard to such data, as a HIPAA Covered Entity. As a Covered Entity, it is able to contract with state agencies and other agencies as a Business Associate to receive identifiable data to perform data matching and de-identification, subject to HIPPA privacy and security requirements that apply to the contracting agency. As a state agency, it is subject to CGSA sec. 4-193.

Recommendation: The Legislature should enact legislation permitting participating agencies to disclose identified information to CHIN, notwithstanding existing state statutes to the contrary, and subject to the privacy and confidentiality requirements of HIPAA.
VI. Governance Plan Recommendations

Structure of CHIN

As CHIN moves toward full implementation, we propose that it consist of two functional entities: a Governing Board comprised of representatives of all entities that contribute data to CHIN, and an Administrative Core which will house and oversee the CHIN infrastructure and operations, to be located at the University of Connecticut Health Center (UCHC). In addition, a broad-based advisory board should be developed, consisting of local directors of health, other community members, researchers and others interested in the availability of health data, to assist in the further development and utilization of CHIN.

Governing Board

CHIN is intended to be a collaborative effort of the participating state agencies and any other entities that may contribute data to the network. CHIN will access data from many databases maintained in several agencies. No one agency can control access to all the databases, since each agency has its own mandates and restrictions on the use and availability of the data it maintains. To this end, the policies, procedures and protocols for the operation of CHIN will be developed collaboratively by the participating agencies. Each participating agency has an identified need for access to the kind of cross-agency data that CHIN will enable, creating incentives to work together to develop mutually beneficial processes and protocols.

Each agency also has differing federal and state privacy and security mandates governing the disclosure of its data, and no one agency can speak to the needs of the other agencies in this regard. While legislative changes by the General Assembly can address state mandates, the state does not have the ability to change federal restrictions. CHIN is being designed to take these differing mandates and restrictions into account, with controls located at each agency to block access from the network to that agency’s data if an individual query would violate any of these mandates or restrictions. These controls were initially developed to allow the US Department of Education to give approval for databases containing educational records to be included in the network, and to be held to be in compliance with FERPA (see discussion below). However, these controls allow all participating agencies to make a determination that a particular requestor should not be permitted access to a particular database for valid reasons, even though access to the databases of other agencies may be permitted. Such determinations will be made by each agency’s representative on the Governing Board, or by such other review entity as the Governing Board may establish for this purpose. It is assumed that the individual agency representatives will be advised in this regard by their agency IRBs and legal counsel, where appropriate.

Recommendation: Establish in statute a CHIN Governing Board to establish policies and procedures for access to data through CHIN. Language from the 2007 proposed legislation is included in Appendix D, and is suggested as a model for such a provision.

Administrative Core

The network itself, including the hardware, software and operations staff, will be administered by the Center for Public Health and Health Policy and located at the UCHC. This is recommended for three reasons: first, the UCHC is seen as a “neutral” site among the other agencies; second, the UCHC is a Covered Entity under HIPAA, which simplifies the Business Associate agreements that need to be developed to comply with HIPAA and ensures that HIPAA privacy and security standards will be met; and third, the UCHC can call on the resources of the University to support the sophisticated technological infrastructure necessary for the further development, implementation, and maintenance of CHIN. The UCHC will maintain a strict separation between the network and its administrative staff on the one hand, and UCHC faculty and researchers on the other hand. No individual
will be allowed to both perform network administrative functions and submit research queries to the network. UCHC researchers will be required to follow the same protocol for gaining access to the network as other agency or outside researchers. (The same will be true for agency personnel who wish to access CHIN.)

Administrative and technical support for CHIN will also be needed at each participating agency or, for agencies whose data are maintained there, at DOIT. The estimated cost of DOIT’s hardware and operating system support, network connectivity, and licensing will be approximately $100,000 per year. It is estimated that, after the initial start-up period, approximately .1-.2 FTE for hardware support will be required at agencies maintaining their own databases. Application support will be the responsibility of UCHC and its contractors.

Recommendation: Establish the Center for Public Health and Health Policy at UCHC as the Administrator of CHIN, with the CHIN Main Application running on network servers located at the UCHC. Establish CHIN Remote Modules at DOIT and at participating agencies who maintain their own databases.

Data Access Procedures

Because the information contained in the participating agencies’ databases are subject to various federal and state privacy and confidentiality laws, it will be necessary for agency personnel and outside researchers who wish to access such data through CHIN to receive approval for their queries from the agencies responsible for maintaining the privacy and security of the data. Applicants will have to demonstrate knowledge of the applicable privacy and security laws, and an understanding and acceptance of their duty to maintain the privacy and security of any data they obtain. Even though CHIN will only release de-identified data in response to approved queries, applicants will be required to comply with the prohibitions against any effort to re-identify individuals from the information obtained.

The application procedures and protocols will be designed by the Governing Board, or a committee designated for such purpose by the Board.

VII. Costs and Funding

Moving forward with the development and production deployment of CHIN will require additional resources beyond that which is currently appropriated. One set of deliverables in our recently executed contract with Akaza Research is a detailed cost estimate for the addition of new agencies and databases to the network. These estimates should be available in June 2009 and will be delivered as an addendum to this report.

Additional funding for CHIN infrastructure development and maintenance may be available through federal sources. Health Information Technology (HIT) funding available through the American Recovery and Reinvestment Act of 2009 (ARRA) should be vigorously pursued through partnerships involving Connecticut’s state agencies and healthcare organizations, as well as the Connecticut General Assembly. In addition, additional monies added to the budgets of the National Institutes of Health via the ARRA present a number of competitive scientific grant opportunities. The Institute for Public Health Research at UCHC is actively pursuing these opportunities in partnership with UCONN’s Department of Computer Sciences and Engineering and Department of Statistics.
References

4. See the following publications:
See the following publications:


Information on DDS databases obtained via conference calls, meetings, and exchange of documentation between Cal Collins & Eddie Tejada of Akaza Research, Rob Aseltine of UCHC, and Mark Warzeca of DDS/DoIT. December 2004 and between R. Aseltine and Laurel Buchanan of UCHC and Tim Deschenes-Desmond of DDS September 2008-February 2009.

Information on DPH databases obtained via conference calls, meetings and exchanges of information between Cal Collins and Eddie Tejada of Akaza Research, Rob Aseltine and Laurel Buchanan of UCHC, and Bruce Wallen, Gary Archambault, Conrad Wopperer, Joseph Lichaa, Lloyd Mueller, and Harinath Chanda of DPH.


Information on Quick Recovery obtained via exchanges of information between Elissa Nagle of UCHC and Robert Aseltine and Laurel Buchanan of UCHC October 2009-March, 2009

Information on NIS obtained via exchange of documentation between Marta Barker and Robert Aseltine and Laurel Buchanan of UCHC October 2009-December, 2008

Information on IDX obtained via exchange of documentation between Sandra Armstrong and Robert Darby of UCHC, Robert Aseltine and Laurel Buchanan of UCHC December, 2008

Information presented by Tim Deschenes-Desmond of DDS and Fred North of DCF at CHIN IT/Program workshop meeting on 7/14/08.

Information on DCF data aggregation and data access improvement efforts obtained from DOIT’s E-Health Project Inventory dated 9/3/2008 and from exchanges of information between Neil Kroke of DCF and Laurel Buchanan of UCHC from December 2008-February, 2009.

Information on DDS data aggregation and data access improvement efforts obtained from DOIT’s E-Health Project Inventory dated 9/3/2008 and from exchanges of information between Tim Deschenes-Desmond of DDS and Laurel Buchanan of UCHC from December 2008-January, 2009.

Information on DPH data aggregation and data access improvement efforts obtained from DOIT’s E-Health Project Inventory dated 9/3/2008 and from exchanges of information between Laurel Buchanan of UCHC and Lloyd Mueller and Marcie Cavacas of DPH December 2008-January, 2009.

Information on OHCA data aggregation and data access improvement efforts obtained from DOIT’s E-Health Project Inventory dated 9/3/2008 and from exchanges of information between Kaila Riggot of OHCA and Laurel Buchanan of UCHC

Information on DHMAS initiatives obtained through exchanges of information between Laurel Buchanan of UCHC and Mark Thomas of DHMAS January, 2009.
Information on DSS initiatives obtained through exchanges of information between Laurel Buchanan of UCHC and Bob Mitchell, Evelyn Dudley, and Jim Wietrak of DHMAS January-February, 2009

Information on state data initiatives obtained from DOIT's E-Health Project Inventory dated 9/3/2008 and from exchanges of information between R. Aseltine and Laurel Buchanan of UCHC

Information on the Data Interoperability Initiative obtained through exchanges of information between Laurel Buchanan of UCHC and Mary Ann Hanley of OWC February, 2009

Information on The Connecticut Data Quality and Access Initiative obtained from James Farnam on 2/9/09
Appendix A
Substitute Bill 7376 (2007 Session), sections 1 and 2
AN ACT ESTABLISHING THE CONNECTICUT HEALTH INFORMATION NETWORK.

Be it enacted by the Senate and House of Representatives in General Assembly convened:
Section 1. (NEW) (Effective October 1, 2007)
(a) There is established at The University of Connecticut Health Center a Connecticut Health Information Network, which shall securely integrate state health and social services data, consistent with state and federal privacy laws, within and across The University of Connecticut Health Center, the Office of Health Care Access and the Departments of Public Health, Mental Retardation and Children and Families. Data from other state agencies may be integrated into the network as funding permits and as permissible under federal law.
(b) The Center for Public Health and Health Policy at The University of Connecticut Health Center, in collaboration with the Departments of Information Technology, Public Health, Mental Retardation, Children and Families and the Office of Health Care Access shall develop, implement and administer the Connecticut Health Information Network.
(c) The Connecticut Health Information Network shall develop a framework for creating the Connecticut Community Health Data and Information Portal, which shall be capable of providing (1) access to public use datasets containing information maintained by state agencies and other nongovernmental entities on the health of Connecticut residents, and (2) a platform to query the network to obtain aggregate data on key health indicators within the state. The Connecticut Community Health Data and Information Portal shall be designed to:
   (A) Provide accurate, timely and accessible health data to public and private sector leaders and policy makers at the state and local level, and inform citizens to improve community and individual health;
   (B) Adhere to strict confidentiality and privacy standards;
   (C) Support efforts to reduce health disparities; and
   (D) Identify the best available data sources and coordinate the compilation of extant health-related data and statistics.

Section 2. (NEW) (Effective October 1, 2007)
(a) There is established a Connecticut Health Information Network Governing Board to oversee the Connecticut Health Information Network established under section 1 of this act.
(b) The governing board shall consist of the following members:
   (1) One appointed by the Governor, who shall serve as the chairperson;
   (2) One appointed by the speaker of the House of Representatives who shall be a local director of health;
   (3) One appointed by the president pro tempore of the Senate who shall be a privacy advocate;
   (4) One appointed by the majority leader of the House of Representatives who shall represent consumers;
   (5) One appointed by the minority leader of the House of Representatives who shall represent data users;
   (6) One appointed by The University of Connecticut Health Center; and
   (7) One each appointed by the Commissioners of Public Health, Mental Retardation, Children and Families and Health Care Access and the Chief Information Officer of the Department of Information Technology, or their designees.
(c) All initial appointments to the board shall be made not later than November 30, 2007. The term of each appointed governing board member shall be four years or until a successor is chosen, whichever is later. Any vacancy shall be filled by the appointing authority.
(d) The chairperson shall schedule the first meeting of the board, which shall be held not later than December 31, 2007.
(e) The governing board shall meet at least once during each calendar quarter and at such other times as the chairperson deems necessary. A majority of the members shall constitute a quorum for the transaction of business.

(f) The duties and responsibilities of the governing board shall be to: (1) Establish and implement policies, procedures and protocols governing access and dissemination of data through the Connecticut Health Information Network; (2) establish such permanent and ad hoc committees as it deems necessary to facilitate the implementation, operation and maintenance of the network; (3) recommend any legislation necessary for implementation, operation and maintenance of the network; (4) perform all necessary functions to facilitate the coordination and integration of the network; and (5) report annually to the Governor and the General Assembly on the status and operations of the Connecticut Health Information Network, including any recommendations for funding.
Appendix B

CHIN Work Group Membership

Policy Work Group

Robert H. Aseltine (UCHC)
Laurel Buchanan (UCHC)
Steve Casey (DOIT)
Melanie Dillon (OHCA)
Mary Eberle (UCHC)
Mark Feller (DCF)
Meg Hooper (DPH)
Joan Twiggs (DCF)
Carolyn Treiss (OHCA)
James Welsh (DDS)

Information Technology and Programmatic Work Group

Robert H. Aseltine (UCHC)
Vasille Buciuman-Coman (Akaza Research)
Laurel Buchanan (UCHC)
Mark Bannon (DOIT)
Steve Casey (DOIT)
Cal Collins (Akaza Research)
Tim Deschenes-Desmond (DDS)
Mary Eberle (UCHC)
Fred North (DCF)
Joan Twiggs (DCF)
Kaila Riggott (OHCA)
Terry Cote (DDS)
Lloyd Mueller (DPH)
Rahman, Aminur (OHCA)
Appendix C

Federal and State Confidentiality Laws and Regulations

I. Federal laws and regulations

Health Insurance Portability and Accountability Act (HIPAA), 42 U.S.C. 201
HIPAA Privacy Rule, 45 C.F.R. parts 160 and 164
Protections for Human Subjects 45 C.F.R. Part 46


Medicare/Medicaid Intermediate Care Facilities for the Mentally Retarded (ICF/MR), 42 CFR part 483.410

II. State laws and regulations

General applicability
CGSA sec. 4-193

DCF
CGSA section 17a-28
CGSA section 17a-101k

DDS
CGSA section 17a-238,
CGSA section 17a-274
CGSA section 45a-668 et seq
CGSA section 52-146c et seq
DDS regulation 19-570-5
DDS regulation 17a-227-1 et seq

DPH
CGSA sec. 19a-2a(10),
CGSA section 19a-7
CGSA section 19a-25
CGSA sections 19a-6e
CGSA section 19a-7h
CGSA section 19a-56b
CGSA sections 19a-41 through 19a-45a
DPH regulations 19a-25-1 through 19a-25-4
CLIA sections 492.1109(e) and 493.2
DPH regulation 19a-36-D32(a)

OHCA
CGSA section 19a-613
CGSA section 19a-654
Appendix D
Substitute Bill 7376 (2007 Session), sections 1 and 2
(Legislation proposed but not enacted in 2007)

AN ACT ESTABLISHING THE CONNECTICUT HEALTH INFORMATION NETWORK.

Be it enacted by the Senate and House of Representatives in General Assembly convened:

Section 1. (NEW) (Effective October 1, 2007)

(a) There is established at The University of Connecticut Health Center a Connecticut Health Information Network, which shall securely integrate state health and social services data, consistent with state and federal privacy laws, within and across The University of Connecticut Health Center, the Office of Health Care Access and the Departments of Public Health, Mental Retardation and Children and Families. Data from other state agencies may be integrated into the network as funding permits and as permissible under federal law.

(b) The Center for Public Health and Health Policy at The University of Connecticut Health Center, in collaboration with the Departments of Information Technology, Public Health, Mental Retardation, Children and Families and the Office of Health Care Access shall develop, implement and administer the Connecticut Health Information Network.

(c) The Connecticut Health Information Network shall develop a framework for creating the Connecticut Community Health Data and Information Portal, which shall be capable of providing (1) access to public use datasets containing information maintained by state agencies and other nongovernmental entities on the health of Connecticut residents, and (2) a platform to query the network to obtain aggregate data on key health indicators within the state. The Connecticut Community Health Data and Information Portal shall be designed to:

(A) Provide accurate, timely and accessible health data to public and private sector leaders and policy makers at the state and local level, and inform citizens to improve community and individual health;

(B) Adhere to strict confidentiality and privacy standards;

(C) Support efforts to reduce health disparities; and

(D) Identify the best available data sources and coordinate the compilation of extant health-related data and statistics.

Section 2. (NEW) (Effective October 1, 2007)

(a) There is established a Connecticut Health Information Network Governing Board to oversee the Connecticut Health Information Network established under section 1 of this act.

(b) The governing board shall consist of the following members:

(1) One appointed by the Governor, who shall serve as the chairperson;

(2) One appointed by the speaker of the House of Representatives who shall be a local director of health;

(3) One appointed by the president pro tempore of the Senate who shall be a privacy advocate;

(4) One appointed by the majority leader of the House of Representatives who shall represent consumers;

(5) One appointed by the minority leader of the House of Representatives who shall represent data users;

(6) One appointed by The University of Connecticut Health Center; and

(7) One each appointed by the Commissioners of Public Health, Mental Retardation, Children and Families and Health Care Access and the Chief Information Officer of the Department of Information Technology, or their designees.

(c) All initial appointments to the board shall be made not later than November 30, 2007. The term of each appointed governing board member shall be four years or until a successor is chosen, whichever is later. Any vacancy shall be filled by the appointing authority.

(d) The chairperson shall schedule the first meeting of the board, which shall be held not later than December 31, 2007.

(e) The governing board shall meet at least once during each calendar quarter and at such other times as the chairperson deems necessary. A majority of the members shall constitute a quorum for the transaction of business.
(f) The duties and responsibilities of the governing board shall be to: (1) Establish and implement policies, procedures and protocols governing access and dissemination of data through the Connecticut Health Information Network; (2) establish such permanent and ad hoc committees as it deems necessary to facilitate the implementation, operation and maintenance of the network; (3) recommend any legislation necessary for implementation, operation and maintenance of the network; (4) perform all necessary functions to facilitate the coordination and integration of the network; and (5) report annually to the Governor and the General Assembly on the status and operations of the Connecticut Health Information Network, including any recommendations for funding.