
Research and Practical Experiences in the Use of Multiple Data Sources for Enterprise-Level Planning and Decision Making: A Literature Review

Using Information in Government Program



**Center for Technology in Government
University at Albany / SUNY**

Executive Summary

The use of multiple data sources for enterprise-level planning and decision making has become increasingly important. Information sharing among organizations can help achieve important public benefits such as increased productivity, improved policy-making, and integrated public services. This paper reviews uses of multiple data sources for enterprise-level planning and decision making. It identifies current research and practical experience in the use of multiple data sources to support performance measurement, strategic planning, and interorganizational business processes. The information was derived from journal articles and Internet sources. A series of cases are examined, and the benefits, issues, methods, and results of efforts that involve the integration of different data sources in the same organization and across multiple organizations are identified and compared. The purpose of this paper is to take the first steps towards the development of a methodology for integrating multiple data sources.

Data integration is the process of the standardization of data definitions and data structures by using a common conceptual schema across a collection of data sources. Integrated data will be consistent and logically compatible in different systems or databases, and can use across time and users. The scope of data integration is the extent to which that the standardization is used across multiple organizations or sub-units of the same organization.

This paper identifies and compares the issues, methods, and results of efforts that involve integrating different data sources 1) within one organization, and 2) across multiple organizations. Section 2) is subdivided into cases (I) where the different organizations are in the same sector of the economy (e.g. in business or government), and (II) where the organizations cross sectors (e.g. business and government).

Integrating data sources within one organization: benefits, barriers, and lessons

In this paper, efforts involving organization-wide data integration are examined in the cases of Kaiser Permanente, Devlin Electronics, Southern Cross, Inc., Greenfields Products, and Burton Trucking Company.

Organization-wide data integration tends to lead to the following benefits in the context of enterprise-level planning and decision making: improved managerial information for organization-wide communication, improved operational coordination across sub-units or divisions of an organization, and improved organization-wide strategic planning and decision making.

However, because multiple sub-units are involved, data integration can also increase costs by increasing the size and complexity of the design problem or increasing the difficulty in getting agreement from all concerned parties. These barriers include compromises in meeting local information needs, bureaucratic delays that reduce local flexibility, and higher up-front costs of information system design and implementation.

Therefore, choosing the appropriate level of data integration in an organization may require trading off improved organization-wide coordination against decreased local flexibility and local effectiveness. In an organization, top management should allow each division to design and implement its own systems, based upon best serving its local needs. Developing a single logical design for use across multiple sub-units can be difficult. Data integration may change the

organizational information flows, and affect individual roles and organizational structure. In addition, the cost of designing and implementing data integration must be considered because it might be much higher than expected.

Integrating data sources across multiple organizations: benefits, barriers, and lessons

Efforts involving data sources integration across multiple organizations are examined in the cases of five states' cancer prevention and control planning model, seven states' statewide cancer control plan, the business process reengineering project at Clark County Recorder's Office, varied databases linking in MassCHIP (Massachusetts Community Health Information Profile), the cancer control intervention in New York State Department of Health, low-income family support at the Child Care Bureau of the U.S. Department of Health and Human Services, assessing hospital performance in QMAS (Quality Measurement Advisory Service) in Washington State, the efforts of the Health Care Data Governing Board in Kansas, and Kentucky's KIDS COUNT Program.

These efforts also tend to lead to the following benefits: increased customers service quality, increased existing personnel efficiency, improved quality, timeliness, and utilization of information, increased accessibility and analysis of information, and the elimination of redundant data and tasks.

While there are clearly advantages, using an integrated approach across multiple organizations presents a number of challenges. Obtaining data from other agencies is often difficult, and in many cases will be impossible. Legal restrictions often prevent access to a particular data set. It is also difficult to obtain the cooperation of agency heads, who will often decide whether to participate in data sharing. Data sharing often requires compatibility between different computer systems as well as the availability of information system personnel. Data integration also requires the concurrence of system administrators, directors of programs, and services consumers. In addition, more cost and time, few data standards, and information overload are also barriers to data integration across multiple organizations.

Based on the experiences in the cases where organizations are either in the same or multiple sectors of the economy, the following important lessons regarding the implementation of a comprehensive data integration project are identified: The purpose of data integration should be very clear. Data integration projects require a significant time commitment. Barriers to participation must be identified and addressed. Early financial commitment is a key to ensuring ongoing political commitment. Management information systems staff should be involved from the start.

It is worth noting that several cases in this paper are in health care field, it probably indicates that health care is a leader in data integration efforts. It also seems to us that organization-wide data integration is done for operational reasons, while data integration across multiple organizations (at least in the cases) is done for research and evaluation purposes.

Introduction

The transformation of numerous and often disparate data sources into knowledge to support critical decisions in a timely manner is essential in today's fast-growing market. Information sharing among organizations can help achieve important public benefits such as increased productivity, improved policy-making, and integrated public services (Dawes, 1996). A large number of disparate data sources are available in organizations, in a variety of formats such as word processing files, flat text files, mail messages, scanned images, spatial data files, audio/voice files, video clips, spreadsheet files, databases, graphics and CAD files. All these data can be derived from different sources either in one organization or across multiple organizations.

The use of multiple data sources for enterprise-level planning and decision making has become increasingly important. In the public sector, it is especially important to integrate the different types and forms of knowledge, and to relate them to the mission of the organization (Dingwall [1], 1998). In today's global economy, enterprises need to better use their information resources to operate more efficiently and effectively. For example, government like business is "in the midst of a trend towards an economy and society based increasingly on knowledge and services" (Dingwall [2], 1998, p14). This requires improved access to timely, accurate, and consistent data that can be easily shared among team members, decision-makers, and business partners (Van Den Hoven, 1998). The collection and organization of information outside an enterprise will become more important and urgent for top management (Drucker, 1998). Enterprise-level planning can help a corporation establish an information system plan to model the primary business sub-systems and applications (Fuhs, 1997). Strategic planning can actively determine the nature or character of the organization and guide its direction. It identifies the mission of the organization and establishes strategies for fulfilling its purposes (Planning Manual, Institutional Research & Planning, 1998). The effectiveness of decision making can be defined by time needed to make decisions, explicitness of decisions, identification and clarification of conflicts, and communication and interpretation of information (Wiggins and French, 1991).

This paper reviews uses of multiple data sources for enterprise-level planning and decision making. It identifies current research and practical experience in the use of multiple data sources to support performance measurement, strategic planning, and interorganizational business processes. The information was derived from journal articles and Internet sources. A series of cases are examined, and the benefits, issues, methods, and results of efforts that involve the integration of different data sources in the same organization and across multiple organizations are identified and compared. The purpose of this paper is to take the first steps towards the development of a methodology for integrating multiple data sources.

The Research Issues: Using Integrated Multiple Data Sources

This research examines the following questions:

- Why integrate multiple data sources?
- What are the benefits of integrating multiple data sources for enterprise-level planning and decision making?

Accessing to the accurate information in a timely manner is a significant challenge facing organizations today. For example, a police officer needs to know if a suspect is wanted in another jurisdiction; A social worker needs to ensure that a welfare applicant is not already receiving benefits elsewhere; A judge needs to see all prior convictions against an offender (McKenna,

1996). These and countless other situations require rapid access to a wide range of complete and accurate information that is often scattered across numerous agencies (McKenna, 1996). However, the problem is that many agencies build information “silos” which are poorly accessible within their own organizations, let alone to the related departments outside the organization. Besides, many agencies seem to have a “genetically encoded political and cultural aversion to information sharing and cooperation, operating instead as isolated fiefdoms or, at best, as grudging partners” (McKenna, 1996).

There has been a spectacular explosion in the quantity of data available in electronic formats in the past few decades. This huge amount of data has been gathered, organized, and stored by a small number of individuals, working for different organizations on varied problems (Subrahmanian, et al., 1996). In light of the ever increasing volume of data, and the expected benefits of integrating the data, a framework for performing integration over multiple data sources is necessary.

What is Data Integration?

Data integration is the process of the standardization of data definitions and data structures by using a common conceptual schema across a collection of data sources (Heimbigner and McLeod, 1985; Litwin, et al., 1990). Integrated data will be consistent and logically compatible in different systems or databases, and can use across time and users (Martin, 1986).

Goodhue et al. (1992, p294) defined data integration as “the use of common field definitions and codes across different parts of an organization”. According to Goodhue, et al. (1992), data integration will increase along one or both of two dimensions: (1) the number of fields with common definitions and codes, or (2) the number of systems or databases adhering to these standards. Data integration is an example of a highly formalized language for describing the events occurring in an organization’s domain. The scope of data integration is the extent to which that formal language is used across multiple organizations or sub-units of the same organization. The objective of data integration is to bring together data from multiple data sources that have relevant information contributing to the achievement of the users’ goals (AFT, 1997).

The Advanced Forest Technologies in Canada (AFT, 1997) identified the following factors which must be addressed to integrate data properly:

- identification of an optimal subset of the available data sources for integration
- estimation of the levels of noise and distortions due to sensory, processing, and environmental conditions when the data are collected
- the spatial resolution, the spectral resolution, and the accuracy of the data
- the formats of the data, the archive systems, and the data storage and retrieval
- the computational efficiency of the integrated data sets to achieve the goals of the users

Benefits of integrating heterogeneous data sources

There are some obvious advantages in integrating information from multiple data sources. Such integration alleviates the burden of duplicating data gathering efforts, and enables the extraction of information that would otherwise be impossible (Subrahmanian, et al., 1996).

Subrahmanian, et al. (1996) gives the following examples of benefits of data integration:

“... law enforcement agencies such as Interpol benefit from the ability to access databases of various national police forces, to assist their effort in fighting international terrorism, drug trafficking, and other criminal activities. Insurance companies, using data from external sources, including other insurance company and police records, can identify possible fraudulent claims. Medical researchers and epidemiologists, with access to records across geographical and ethnic

boundaries, are in a better position to predict the progression of certain diseases. In each case, the information extracted from the integrated sources is not possible when the data sources are viewed in isolation.”

Integrating diverse data source paradigms

Subrahmanian, et al. (1996) established a data source paradigm. There are two important aspects to constructing the data source paradigm: domain integration and semantic integration. Domain integration is the physical linking of data sources and systems, while semantic integration is the coherent extraction and combination of the information provided by the data and reasoning sources, to support a specific purpose (Subrahmanian, et al., 1996).

It is acknowledged that data warehousing is the most effective way to provide the business decision support data (Van Den Hoven, 1998). Under this concept, data is derived from operational systems and external information providers, and subsequently conditioned, integrated, and changed into a read-only database that is optimized for direct access by decision makers. The term ‘data warehousing’ describes data as an enterprise asset that must be identified, cataloged, and stored to ensure that users will always be able to find the needed information. The data warehouse is generally enterprise-wide in scope, and its purpose is to provide a single, integrated view of the enterprise’s data, spanning all enterprise activities (Van Den Hoven, 1998).

This paper identifies and compares the issues, methods, and results of efforts that involve integrating different data sources 1) within one organization, and 2) across multiple organizations.

Integration Efforts Involving Different Data Sources within One Organization

a. Kaiser Permanente: Integrating Legacy Systems (Robertson, 1997)

Kaiser Permanente, a health care service provider, used three related technologies (dynamic OOP, domain-specific embedded languages, and reflection) in one project integrating their legacy systems. Typically, a legacy system is an in-place structure that is neither optimal for modern needs nor modifiable for project purposes (Robertson, 1997). The data often does not reside in a single database or in a single format. More often, it is distributed across a number of different vendor databases, running on different platforms, with significant physical distances between the separate services. It is often needed to use a variety of data sources for report generation, management information system construction, or the creation of client-server or Intranet-based applications. The Kaiser Permanente legacy systems contain data such as membership, subscription information, pharmacy, drugs, appointments and encounters, and billing. The data come from a variety of sources including online connections to pharmacies, and data input forms completed in doctors’ offices by doctors and patients. The accuracy of the data is critical for accurate billing, accurate payments to service providers including consultants, physicians, and pharmacies, and the establishment of appropriate member status for a patient.

There are many areas in which legacy data can be put to use, such as marketing, executive, government reporting, competitive analysis, and new access to data. Legacy data contains a wealth of information that can be used in promoting a company’s products, in running a business efficiently, and in providing competitive services. There are many opportunities for using legacy data, but the pressure to take advantage of legacy data is extremely high (Robertson, 1997).

b. Devlin Electronics: Tracking Production Schedules (Goodhue et al., 1992)

Goodhue et al. (1992) analyze the case in Devlin Electronics: when on-time deliveries in Devlin Electronics fell to only 70 percent, a multi-disciplinary team used organization-wide integrated scheduling data to track how production schedules were developed, changed, and adjusted by the different groups involved. They found a number of interrelated problems such as not properly updating inventory levels and equipment conditions at some plants, schedule overridden by marketing, neglecting plant capabilities and critical order requirements, etc.. By using organization-wide integrated scheduling data, Devlin Electronics understood its problems and then took corrective action. As a result, on-time delivery increased from 70 percent to 98 percent.

c. Southern Cross: Cross-product Line Analysis (Goodhue et al., 1992)

When the executive vice president (EVP) of Southern Cross, Inc. needed to find out why sales were up by only one-half percent in spite of many new accounts, and demanded an analysis across all regions, customers, and products to determine the cause, nothing in the standard reports provided any indication of the problem. Because the information was contained on several different (unintegrated) systems, and the products had been grouped into various categories, there was no way to conduct an automated analysis. After 40 person-hours of effort of using spread-sheet programs and manually backing out products that had switched categories and adjusting inconsistencies between the different systems, the top-level analysts assembled a compatible base of information to answer the EVP's questions. Their analysis indicated that the sales slump was occurring primarily in the old, established, long-time distributorships (an insight that was not apparent from analysis of the non-integrated data). With the problem pinpointed, managers could take appropriate further action (Goodhue, et al., 1992).

d. Greenfields Products: Creating A Single Customer Interface (Goodhue et al., 1992)

Goodhue et al (1992, p301) analyzed the case of Greenfields Products:

“Each of Greenfields Products’ five divisions has its own salespersons and distributes its own product lines in separate trucks. Top management wanted the ability to coordinate the sale and delivery in the five divisions, that is, to present a single face to the customer by having only one salesperson (not five) call on each customer and only one truck (not five) back up to the customer’s delivery dock. They realized that without an integrated, consistent base of customer and order data, coordinating the actions of the five divisions to create a single customer interface would be impossible.”

e. Burton Trucking Company: Better Dispatching and Shipment Tracking (Goodhue et al., 1992)

Burton Trucking Company uses its information systems based on a single integrated data model for the entire company, where the integrated data was derived from each sub-unit. This integrated system allows them to link across both geography and functions. By using integrated, sharable data, they expanded their dispatch systems (the responsibility of operations) with little effort to have a much better shipment tracking system (the responsibility of marketing). Data integration made them capitalize on previously unrecognized interdependencies between dispatching and shipment tracking.

Though the new dispatching system “contained integrated data about all customers, equipment, and shipments, salespeople at each local terminal argued that in order for the information to be valuable to them, they needed to add additional fields such as permissible delivery hours, after-hours phone numbers, and special instructions for drivers. But the salespeople could not agree on exactly which additional fields should be added. Terminals with close-in satellites (trucks every 2-3 hours) had very different needs from those with distant satellites (trucks every 5 hours). It was decided that trying to standardize at this level did not make sense. They designed 10 extra fields that the local

people could use as they saw fit and gave them search capabilities and screens to update and query whatever data they needed in those fields” (Goodhue, et al., 1992, p302).

Also at Burton Trucking Company, the operations group wanted to automate freight transfer recording using bar codes, but because the data administration group did not understand what they were trying to do so could not “square” it with their data model. After a delay of six months, the project could proceed and was quite successful (Goodhue, et al., 1992). This case indicates that the requests for change in the use of integrated data may involve certain bureaucratic delays.

Integration Efforts Involving Different Data Sources across Multiple Organizations

According to Drucker (1998), in the next 10 to 15 years, collecting information from external sources will be the next frontier. Following are cases of research or practical experiences in the use of multiple data sources across multiple organizations. They are subdivided into cases (I) where the different organizations are in the same sector of the economy (e.g. in business or government), and (II) where the organizations cross sectors (e.g. business and government).

I. Cases of organizations in the same sector of the economy

1a. Five States’ Cancer Prevention and Control Planning Model (Alciati and Glanz, 1996)

Alciati and Glanz (1996) described the results of an analysis of five states’ experiences using and integrating available data to develop cancer control plans for their states. These states included Georgia, Maryland, North Dakota, Vermont, and Washington State. In 1989, the National Cancer Institute funded the second round of Data-Based Intervention Research (DBIR) cooperative agreements with state health agencies to implement a four-phase planning model to establish ongoing cancer prevention and control programs. Activities focused on the identification and analysis of data relevant to the development of a state cancer control plan. Alciati and Glanz’s research explores how states use different types of available data to make public health planning decisions, the levels of sufficiency of data for this planning, and the perceived costs and benefits of a data-based planning approach.

According to Alciati and Glanz, while using health data to guide public health planning efforts is not new, information on how states use existing multiple data sources for comprehensive cancer prevention and control planning is limited. What is lacking is “a clear picture of how these components fit together in a comprehensive state-level planning process, how data are used to establish cancer prevention and control priorities and to identify proven interventions for implementation, and what states perceived to be the costs and benefits of such detailed data-based planning.”

Data sources and integration methods

Each state used three categories of data:

- *Health Data* – including mortality and morbidity (incidence); states generally relied on a small number of measures, such as the number of state deaths, age-adjusted death rates for the state, and survival.
- *Behavioral Data* – including health behavior, risk factors, and determinants of behavior (for example, knowledge, attitudes, and beliefs). The Behavioral Risk Factor Surveillance System (BRFSS) was the primary source of state-specific behavioral data. Behavioral data were used primarily to identify target groups for intervention.
- *Environmental and Health Services Data* – including environmental characteristics such as the presence of cancer control legislation and worksite policies, the availability of early

detection equipment to support public health goals in cancer prevention and control, as well as information about the existence of cancer control programs and the utilization of health services. The most important sources for this information were hospital discharge datasets and state and local surveys.

For each type of data, the specific data source, the measures used, the type of subgroup analyses performed, and to the extent possible, how the data were used to establish planning priorities and identify interventions were recorded in this research. The data was then summarized, to identify the number of states using each type of data source, data measure, and subgroup analysis as well as the number of states using data to make each type of planning decision.

In these five state programs, comprehensive cancer control planning efforts used a full range of integrated data, and linked these data to decision-making for cancer control. “This research also provides a framework for public health planners to identify the type of data likely to be available for cancer prevention and control planning at the state level, various measures that can be realistically derived from these data, and how they can be linked to public health planning” (Alciati and Glanz, 1996).

1b. Seven States’ Health Department: Developing a Statewide Cancer Control Plan (Boss and Suarez, 1990)

Seven state health departments in Illinois, Nebraska, New Jersey, New York, North Carolina, Texas, and Wisconsin, participated in an effort to utilize a variety of state-specific cancer-related data to describe the cancer burden in their state’s population. The data were then used to develop a statewide cancer plan or to supplement an existing plan to address the defined problems. The efforts in these states can serve as models for data use to prevent and control cancer and other chronic diseases. State-specific data can be used to rank needs and make a clear case that can influence resource allocation decisions. In this research, Boss and Suarez described the data sources and additional statistics that were used to provide a broad picture of the cancer burden which can assist in targeting and defining intervention needs.

The Problem

According to Boss and Suarez, more data exist to describe cancer than any other disease. However, the data have been rarely systematically evaluated to target and plan public health programs in cancer prevention and control. Program planning has often been based on historical or political priorities, and therefore programs have not necessarily been located where the need or potential impact is the greatest.

Potential Solutions and Data Sources

Four major data sets were used by these states: 1) mortality data, 2) incidence data, 3) risk factor data, and 4) hospital discharge data. These data sets appear to be the most accessible and potentially useful of the examined data sources. Various additional data sets were also used, many of which are available within state government, and often within health departments. Data sources used to describe the facilities within the state included the ACOS listing of hospitals with approved cancer programs and information from the local officials of the American Cancer Society (ACS), Cancer Information Service, and the radiologic health unit of the state health department. Information on personnel resources came from state medical organizations and the state boards of medical examiners. Environmental data bases included lists of abandoned landfills and results of water and air monitoring. Limited treatment information could be obtained from ACOS patterns of care surveys, cancer centers, and public health clinics. State taxation records provided information on cigarette and smokeless tobacco sales and tax income.

Other sources were also used. Various types of insurance claims data were examined, such as Medicare, Medicaid, Blue Cross, and State Employees Insurance records. National data sources were also used primarily for comparison with local data. The national sources included the National Health and Nutrition Examination Survey (NHANES), Hispanic Health and Nutrition Examination Survey (HHANES), and Nationwide Food Consumption Survey. For example, the Texas sample from HHANES was large enough to provide state-specific data on the Hispanic population. New data variables can be derived from a variety of existing information sources. Such information is essential to plan programs that meet a particular goal.

The above two research cases focused on using different data sources in organizations in the same sector of the economy, the government sector.

II. Cases of organizations in multiple sectors of the economy

2a. Clark County Recorder's Office: Business Process Reengineering (<http://www.co.clark.nv.us/recorder/>)

The Recorder's Office in Clark County "ensures the timely recording, microfilming, and permanent retention of documents; keeps an accurate and accessible indexing system for the research and retrieval of recorded documents; and ensures fiscal responsibility for the funds and fees collected while providing a cost-effective and efficient service to customers."

The Recorder's Office is responsible for maintaining the public record of Clark County, Nevada. The Recorder has maintained all real estate transactions, financing documents, maps, mining claims, military papers, declarations of homestead, mechanic's liens, marriage certificates, and real property transfer tax. These data sources are integrated from different organizations across government and business sectors.

Over the years, the method of recording documents has evolved from a system of manual transcription to microfilming and computerized indexing of documents. In order to respond to the critical information needs of customers, the business process reengineering (BPR) team implemented the reengineering of processes along the development of an integrated system to improve efficiency and accuracy in the recording process, and to provide increased quality of customer service. The driving force behind the BPR project is customer satisfaction. The results of the project can be useful for the county to develop standards for technological improvements on customer service.

2b. SEI's MassCHIP: Linking Varied Databases (<http://commonwealth2.Cam-colo.bbnplanet.com/dpd/dphhome.htm>)

At Software Engineering Institute (SEI) of Carnegie Mellon University, its MassCHIP (the Massachusetts Community Health Information Profile) links varied databases for the Massachusetts Department of Public Health. Statistical public health data is collected by a number of public and private organizations within the Massachusetts Department of Public Health (DPH). The DPH is responsible for health service planning and relies heavily on these diverse sources for accurate information.

Problems and Barriers:

Prior to working with SEI, the focus and structure of the data in DPH were not coordinated, limiting the benefit of gathering the great expanse of information. The problems were: multiple

agencies sharing information manually; 37 separate databases existed within the DPH, unconnected and user unfriendly; data existing on many platforms and in many incompatible formats; data accessing was difficult, time-consuming, and consequently expensive; inability to query data from separate databases; a single means of extracting and storing data was needed to streamline the transfer of information between agencies. Technically, it was necessary for the database structure to accommodate a variety of data types. In addition, a query navigator was required so that inexperienced users could build queries. A mechanism for presenting selections of data in a variety of ways to a variety of viewers was also needed.

Solutions:

DPH worked with SEI to address the above problems. SEI developed a classic data warehouse for the health and statistical information provided by many state agencies and some private concerns. The system is called MassCHIP (Massachusetts Community Health Information Profile). It is an infrastructure that allows data sharing between departments. The user can view the data by using tabular formats, a variety of charts, and geographical maps, and data may also be exported into many formats for uses with analysis tools of the user's choice. The various sources of data use different coding schemes to store information. Platform independence and the ability to query across multiple databases have been realized. Centralized directory services are supported and data become available to anyone with Internet access. In addition, a data transformation utility was developed to translate multiple sources of data into the MassCHIP structure.

This is an example of a case of where government agency collaborated with business and technical professionals to support health service planning.

2c. New York State Department of Health: Cancer Control Intervention (Lillquist, et al., 1994)

A number of data sources routinely available to state health departments can be analyzed as part of a state health department cancer control planning effort. In 1986, the National Cancer Institute (NCI) initiated the Data-based Intervention Research (DBIR) Program, a program of grants and cooperative agreements awarded to state health departments to build ongoing cancer control programs to ensure the translation of cancer prevention and treatment science into practice. The New York State Department of Health (DOH) was one of the first six states funded under this Program in 1987.

New York State DOH efforts involved the following steps:

- *Identifying Data* -- a total of 27 different sources of data were identified for evaluation. In general, these were population-based and included information on all New York State residents.
- *Defining Data Characteristics*
- *Assessing Data Usefulness and Quality*
- *Analyzing Data*
- *Defining the Cancer Burden*
- *Prioritizing the Cancer Burden*

Data sources of population-and non-population-based data for New York State include:

- New York State Department of Health: Cancer registry, Statewide Planning and Research Cooperative Data System, Family planning data system, Heavy metals registry, Vital records mortality data, and Behavioral Risk Factor Surveillance System
- New York State Department of Taxation and Finance: Cigarette tax information

- New York State Department of Environmental Conservation: Industrial chemical survey
- Roswell Park Cancer Institute(RPCI): tumor registry of RPCI patients, patient epidemiology data system, patterns of care data base
- U.S. Department of Commerce: U.S. census data

Assessment of the magnitude of the problem was based on five additional factors:

- Impact of the cancer on the population as a whole
- The impact of the cancer on specific sub-populations
- Impact of the cancer on the medical care system
- Time trends in incidence
- Risk factor prevalence

The experience of the New York State DOH suggests that “state and local health departments have access to data sources that are useful in cancer control planning and the establishment of priorities for public health action. Combined with a systematic approach to planning, these data provide a solid foundation for ensuring that limited resources are directed to areas of greatest need and support efforts with the highest probabilities of success”(Lillquist, et al., 1994).

The application of this planning process and framework for setting intervention priorities in New York State also revealed several other important facts (Lillquist, et al., 1994): 1) data were unavailable for a number of cancer control areas that may otherwise have been chosen for intervention. “Work on this project enhanced recognition of the lack of information in some priority areas and stimulated developments to collect it”; 2) the data that were available were most useful in assessing the impact of various forms of cancer on the population and in identifying sub-populations with unusually high rates of disease or exposures to known risk factors. These data provide the foundation for targeting intervention efforts; 3) assessment of information specific to local communities and target groups was important for several reasons, such as providing the foundation for evaluating intervention outcomes; and 4) the planning process and framework used by New York might be useful for similar efforts in other states.

2d. Child Care Bureau: Supporting Low-Income Families (<http://www.acf.dhhs.gov/programs/ccb/data/index.htm>)

The purpose of the Child Care Policy Research Consortium is “to examine child care as an essential support to low-income families in achieving economic self-sufficiency while balancing the competing demands of work and family life”. Research partnerships funded by the Child Care Bureau include state child care agencies, university research teams, national, state and local child care resource and referral networks, providers and parents, professional organizations, and businesses.

The Administration for Children and Families’ (ACF) child care programs are administered by the Child Care Bureau within the Administration on Children, Youth and Families in the U.S. Department of Health and Human Services (DHHS). ACF’s child care programs help low-income families to get child care and other supportive services so they can work or participate in an approved education and training program, then achieve economic self-sufficiency.

The Child Care Bureau has created a Child Care Automation Resource Center to provide better service delivery and support for child care programs; and to ensure more reliable data collection on all child care recipients and providers and timely, accurate child care data reporting.

In addition to the practice in child care at the Child Care Bureau, issues such as the value and shortcomings of state and local administrative databases as a source of data on child care were also addressed (Approaches to Data Collection, Chapter 5, 1997). The promising opportunities to forge partnerships between academic researchers and federal and state child care agencies, resource and referral agencies, and others who manage local databases has been noted. The need for efforts aimed at improving the comparability of data across local and state databases was also addressed.

2e. QMAS: Assessing Hospital Performance (<http://www.qmas.org/tools/guide-assessing/33total.htm>)

The Quality Measurement Advisory Service (QMAS) report introduces “the range of dimensions of hospital quality that can be measured, reviews some of the diverse tools that are available today, and presents illustrations of how they are being applied in the ‘real world’ by various organizations and associations around the country.” The report is based on the proceedings of a workshop sponsored by the QMAS in the spring of 1997.

The QMAS was established in Washington State in early 1996 to assist state and local health care coalitions, purchasing groups, and health information organizations in their efforts to measure health care quality. It is a not-for-profit collaborative initiative of the Foundation for Health Care Quality, the Institute for Health Policy Solutions, and the National Business Coalition on Health.

Measuring the dimensions of hospital quality is clearly a complex and challenging task. There are three basic sources of data used in QMAS hospital assessment: 1) administrative files (e.g., claims or bills), 2) medical records, and 3) patient survey results. To determine whether a data source will be feasible and adequate to the task, it is critical to determine the information it contains, its accuracy and reliability, which patients are included, the cost, whether the data are computer-readable, and the currency of the data.

Using Administrative Data

Administrative data refer to information generated as a by-product of administering care and services in the hospital, primarily from billing for reimbursement or from efforts to meet regulatory requirements. The data typically contain information such as patient demographics, diagnostic codes and procedures performed, and the charges billed to payers. Hospital administrative data are typically drawn from the following three sources: 1) payers, 2) state health data organizations, and 3) hospital associations. Administrative data are computerized and easy to analyze. They are also typically inexpensive and can be used to assess quality, efficiency, and other performance issues.

Using Patient Surveys

While administrative data offer information on care from the hospital’s perspective, the patients’ perspectives can also play an important role in the measurement of health care quality.

Patient surveys are important in assessing quality because they are useful for determining the patients’ viewpoints about the care that they received; they assess the quality of interpersonal communications, and the patients’ physical and psychosocial functioning outcomes. Surveys can identify problems and actions depending upon the nature of the questions asked. The rich data provided by surveys can pinpoint specific problems in the delivery process, and directly identify actions to improve care.

Using Clinical Data

Clinical data refer to the clinical attributes of patients, and represent factors that health care professionals use for patients, such as symptoms (e.g. chest pain), vital signs (e.g. blood pressure),

and laboratory test results. They are the types of observations written down by health care providers in the medical record, and are the data used to diagnose patients and determine treatment plans. Clinical data are derived primarily from medical records, which contain detailed information about each patient. Clinical data capture extensive dimensions of quality. They address a broader array of quality dimensions than can be addressed with administrative data. To be useful for quality measurement purposes, clinical data must be made computer-readable, which can be costly and time-consuming.

Clinical data also have some drawbacks, such as: 1) expensive cost of data collection. Electronic medical records that may reduce the cost of data collection are not yet used widely; and 2) lack of integration with outpatient and preadmission information. Crucial information remains in the patients' outpatient records in their doctors' offices, and hospital and physician patient records are usually kept separate.

The following three practices illustrate some of the organizational strategies that have been adopted in different geographic areas.

Practices in Cleveland, Ohio

Cleveland's quality measurement program was initiated in the early 1990's by business leaders associated with the Health Action Council of Northeast Ohio, an alliance of 140 large employers. With help from Cleveland Health Quality Choice (CHQC), an organization established for this purpose, the group examined clinical quality and patient satisfaction with Cleveland-area hospitals. CHQC has published biannual reports that hospitals can use in quality improvement programs since 1993. Some Council members also use the data in their purchasing decisions.

The project to measure hospital quality in Cleveland started with two strong assumptions: 1) employers would reward hospitals that demonstrate good value, based on the CHQC data; and 2) the providers would use the data to improve service delivery. Based on their experiences using an integrated data set to assess quality and decision making, they offered the following advice (<http://www.qmas.org/tools/guide-assessing/33total.htm>):

- Disregard disclaimers about the data quality because the collected data is always useful in some fashion
- Solicit hospital participation directly. Hospital associations may oppose measures that might jeopardize an individual facility's status
- Carefully consider which parties should be involved at the beginning
- Consensus between providers and purchasers is very important to a quality-driven health care marketplace
- Coalitions have important roles in consensus building and their role will become very political in nature
- Experienced vendors should be selected for quality measurement initiatives
- Create a technical advisory board, including members from the business community, information systems professionals, statisticians, and physicians
- Share data with the public, but this can change the dynamics of the relationship between purchasers and providers, and limit meaningful dialogue on less rigorously obtained data
- Realize that the data will be used in ways that were not anticipated
- Data can become stale if not used to drill deeper into quality performance issues

Practices in St. Louis, Missouri

The Greater St. Louis Health Care Alliance is a voluntary organization of 22 hospitals, 7 managed care organizations, and 30 business members. Hospitals are evaluated by the measures of risk-adjusted clinical outcomes and patient satisfaction. The Alliance chose to gather information on these measures from four vendors (<http://www.qmas.org/tools/guide-assessing/33total.htm>):

- The Picker Institute -- in St. Louis, Picker collects patient satisfaction survey data from 12,000 telephone interviews each year
- Aspen Systems Corporation -- Aspen abstracts clinical data from medical records
- Michael Pine & Associates -- they report clinical data that have been adjusted according to physician-reviewed models that use statistically significant risk factors
- Cleveland Health Quality Choice -- CHQC analyzes the data and produces the comprehensive and summary reports

Mr. Sutter, the program director, offers the following lessons for those who are considering an integrated approach to quality measurement (<http://www.qmas.org/tools/guide-assessing/33total.htm>):

Managerial lessons:

- Garner genuine collaboration and commitment up front
- Specifically define goals at the beginning of the project
- Provide extensive education for users
- Select experienced, qualified vendors and draft explicit contracts
- Conduct a timely, effective quality review process
- Use project management techniques to manage schedules, deadlines, and logistics

Measurement lessons:

- Do not measure hospital costs. It is not worth the time, money, or effort to measure them
- Determine how data for transferred patients and patients discharged to skilled nursing facilities will be handled
- Quantify the appropriate sample size to ensure statistical validity
- Provide a combined performance measurement, reflecting both clinical outcomes and patient satisfaction results

Practices in Madison, Wisconsin

The Employer Health Care Alliance Cooperative, known as the Alliance, was founded in 1990. It is an employer-owned, nonprofit cooperative based in Madison, Wisconsin. In 1997, the Alliance represented nearly 120 large to mid-size employers with approximately 25,000 employees, as well as over 700 small employers with about 7,000 employees. The Alliance's interest in quality measurement data is twofold: 1) to improve the health delivery services and the health status of its members' employees; 2) to support the vision of accessible, affordable and effective care, three integrated strategies that drive quality improvements are employed: collaboration, informed self-interest, and contractual incentives. The three strategies are described below (<http://www.qmas.org/tools/guide-assessing/33total.htm>):

Collaboration

Collaboration, defined as two or more organizations working together to achieve a common goal, is appropriate for situations where all parties agree that a project is worthwhile, yet beyond the scope of any one organization. While collaboration can be useful in such situations, for fostering learning and relationship-building as well as serving as a motivating force to attract others' involvement, it can also be very time consuming and resource intensive.

Informed Self-interest

Informed self-interest is a strategy “where comparative data is shared privately or publicly with providers and there is a commitment to monitor the data on a periodic basis.” This strategy is useful in situations where variability exists among organizations. This strategy can create expedient results and foster competition, but it can also engender defensiveness from hospitals. In addition, it may undermine the measurement if the project focuses intensely on one or two problem areas.

Contractual Incentives

Contractual incentives involve the use of performance targets in determining hospital reimbursement. Structuring reimbursement to reflect performance is useful when other approaches are not working. It is important that purchasers are willing to pay for it, and results in purchasers having access to valid and reliable data to substantiate hospitals’ performance.

2f. Health Care Data Governing Board in Kansas (<http://www.ink.org/public/hcdgb/khcd95report.html>)

A lack of standardized data among various sources is one of the most serious barriers to sound health care needs identification and decision making. In Kansas, health care occupations information is maintained within numerous agencies, and a series of discussions were held with the Kansas health care occupations credentialing boards to discuss a minimum data set for health occupations in Kansas. A minimum dataset was recommended based upon recommendations from the National Center for Health Statistics (NCHS). The Health Care Data Governing Board developed policies for collection, security analysis of health data, and dissemination of information from the Kansas health care database. In 1995, the Governing Board implemented its first data collection initiative, a health system inventory. Subsequently, centralized and evaluated data was collected on health-care occupations from eight credentialing boards encompassing 23 occupations. The mechanisms bring data issues to the forefront, and data products are available for decision making.

The accomplishments of Kansas’ health care database developers and partners include (<http://www.ink.org/public/hcdgb/khcd95report.html>):

- Publication of the Kansas Health Data Resources Directory which provides a concise guide to health data collected within the state
- Development of centralized computer and analytic resources within Kansas that standardize, house, analyze, and disseminate health data
- Implementation of the health system inventory and publication of the first health care provider standard reports which are tabulations of health care occupations within Kansas
- Development and approval of a model data collection instrument for health care occupations to serve as a guide for data collection on these occupations in Kansas
- Acquisition of hospital discharge summary data for analysis and public distribution
- Development and consensus on health status indicators to be collected for the database using sentinel measurements to monitor the health status of Kansans
- Revision of policy questions to increase the priority for evaluating the quality of health care in Kansas

The Kansas Health Data Resource Directory catalogs the health data resources maintained in Kansas state government, universities, and private agencies. It serves as a pointer to locate health information collected in Kansas. It also provides information about who may be contacted about the data. The Resource Directory serves as a reference about the kinds of data available, helps to

identify duplication in data collection, and ultimately may facilitate data sharing between agencies. It will also be useful to anyone collecting or using health care data.

Potential Policy Questions Needing Data Support

A lot of policy questions were raised by the Health Care Data Governing Board to ensure that the data collected and maintained in the health care database are relevant. These potential policy questions can be used to guide the database design. The questions relate to the availability and distribution of health care services, utilization, expenditures, health status, and outcomes. To answer these questions, the following need to be identified and determined with immediate priority (<http://www.ink.org/public/hcdgb/khcd95report.html>):

- the providers available in the state and the services they are providing
- the sources and applications of funding for health services by provider and payer type
- the demographic characteristics of the uninsured and underinsured
- the distribution and access problems for health services in Kansas
- the utilization of health services in Kansas
- the characteristics of the population utilizing public health services
- the health status of Kansans
- the cost to insure the uninsured and underinsured population
- the differences of utilization patterns and resulting outcomes across Kansas
- the services provided by the primary care providers in Kansas
- the portion of the health care dollar spent on preventive medicine
- the services provided by public health departments
- the full costs of medical litigation
- the effectiveness of operating service networks developed under health care reform
- the effects of Kansas risk adjustment factors on community rating
- the effects of insurance mandates on premium costs
- the results of cost and outcome comparison between and among types of primary care professionals
- the utilization rates and costs of common procedures for individual hospitals, clinics, ambulatory centers and community health centers
- the impact of health care reform on quality of care

An integrated data system of health care providers is being developed. It will allow users to analyze data across professions and facilities from multiple data sources. The data will be available to customers via: 1) Internet access and electronic data transfer; 2) information formats generated through standard reports and special requests; and 3) publications and media articles (<http://www.ink.org/public/hcdgb/khcd95report.html>).

2g. Kentucky KIDS COUNT: Affecting Public Policy on Welfare Reform (<http://www.louisville.edu/cbpa/kpr/kidscount/define96.htm>)

The Kentucky KIDS COUNT is “a unique consortium of researchers and children’s activists who have significant expertise in the aggregation, interpretation, and use of data to affect public policy.” The consortium’s work includes producing a series of reports on children and families to publicize the needs of children, influencing budget and programs decisions, and monitoring state and local performance for children.

Multiple data sources are used for the 1996 Kentucky KIDS COUNT profiles of the status of children and their families in the state and its 120 counties and for the profiles of education-related data in the 176 local school districts in the state. For example, data on children’s enrollment rates,

attendance rates, dropout rates, free and reduced lunches, and per pupil spending were provided by the Kentucky Department of Education. Data on child abuse and neglect were derived from annual data compiled by the Department of Social Services, and the Cabinet for Families and Children. Data on food stamps were provided by the Department of Social Insurance, and the Cabinet for Families and Children. Data on Medicaid were derived from data compiled by the Department for Medicaid, the Cabinet for Families and Children, and the Social Security Administration. Population data were obtained from the Urban Studies Institute, the University of Louisville, and population forecasts. Data on pupil/teacher ratio, retention rates, and revenue by source were provided by the Kentucky Department of Education, Division of Finance. Data on supplemental security income (SSI) were obtained from data compiled by the U.S. Social Security Administration.

A Summary of Integrating Information from Diverse Data Sets

(1) Efforts involving organization-wide data integration: benefits, barriers, and lessons

Benefits

Organization-wide data integration tends to lead to the following benefits in the context of enterprise-level planning and decision making:

- Improved managerial information for organization-wide communication (Goodhue, et al., 1992)
- Improved operational coordination across sub-units or divisions of an organization (Goodhue, et al., 1992)
- Improved organization-wide strategic planning and decision making

Data integration is necessary for data to serve as a common language for communication within an organization. Without data integration there will be increased processing costs and ambiguity between sub-units or divisions. Without data integration, there will be delays and decreased levels of communication, reductions in the amount of summarization, and greater distortion of meaning (Huber, 1982). Data integration facilitates the collection, comparison, and aggregation of data from various parts of an organization, leading to better understanding (Goodhue, et al., 1992), and improved enterprise-level planning and decision making when there are complex, interdependent problems.

Barriers

Data integration can have a positive impact on reducing costs by reducing redundant design efforts (Goodhue, et al., 1992). However, because multiple sub-units or divisions are involved, data integration can also increase costs by increasing the size and complexity of the design problem or increasing the difficulty in getting agreement from all concerned parties.

These barriers were summarized by Goodhue et al. (1992) as:

- Compromises in meeting local information needs
- Bureaucratic delays that reduce local flexibility
- Higher up-front costs of information system design and implementation

Organization-wide data integration may result in a loss of local autonomy in the design and use of data. In addition, it may also involve a loss of local effectiveness. Over time, different sub-units may face different task complexity and environmental challenges of unanticipated local events (Sheth and Larson, 1990).

Lessons

The following lessons have been derived from the cases of organization-wide data integration (Goodhue, et al., 1992; Sheth and Larson, 1990; Robertson, 1997):

- Choosing the appropriate level of data integration in an organization may require trading off improved organization-wide communication and coordination against decreased local flexibility and effectiveness
- The top management in an organization should allow each division to design and implement its own information systems, based upon best serving its local operational and information needs. “The result might be systems that are locally optimal but not integrated across the divisions, with different definitions, identifiers, and calculations in each division” (Goodhue, et al., 1992, p302)
- A single logical design for use across multiple sub-units can be difficult. The more sub-units involved and the more heterogeneous their needs, the more difficult it will be to develop a single design to meet all needs
- Data integration may change the organizational information flows, and affect individual roles and organizational structure
- The cost of designing and implementing data integration must be also considered, because it might be much higher than expected

(2) Efforts involving data integration across multiple organizations: benefits, barriers, and lessons

Benefits

Like the efforts involved in single organization, the use of multiple data sources provide improved communications and coordination across different organizations, both in the same and different sectors of the economy. In addition, these efforts involved in the cases also tend to lead to the following benefits (Clark County Recorder’s Office, 1998; QMAS Report, 1997; SEI’s MassCHIP Program):

- Increased customers service quality
- Increased existing personnel efficiency
- Improved information quality, timeliness, and utilization
- Increased accessibility, and analysis of information
- The elimination of redundant data and tasks

Barriers

While there are clearly advantages, using an integrated approach across multiple organizations also presents a number of challenges. For example, obtaining data from other agencies is often difficult, and in many cases will be impossible. Culhane and Metraux (1998) summarized these limitations as:

- Legal restrictions may prevent access to a particular data set
- Difficulty in obtaining the cooperation of agency heads, who will often make data sharing decisions based upon “perceived self-interest for the agency or the current political administration”
- Data sharing often requires compatibility between different computer systems as well as the availability of information system personnel with the requisite time and technical skills
- Integrating data systems frequently requires the concurrence of system administrators, directors of programs, and services consumers

Other barriers are also identified from the cases discussed above (QMAS Report, 1997):

- *Cost* -- Data integration across organizations can cause expenses to multiply. For example, conducting multiple performance evaluations simultaneously may be more expensive than using each tool separately
- *Timing* -- It takes much more time to collect needed data from different sources across organizations. This time lag may cause synchronization problems
- *few data standards* -- This will result in no clear vision of data strategy, and make the information decision support much more complex
- *Information overload* -- Organization staff may be overcome by the volume of information across multiple organizations. They may view this abundance of information as overly complicated, and may choose not to use it

Lessons

Based on the experiences in the cases where organizations are either in the same or multiple sectors of the economy, the following important lessons regarding the implementation of a comprehensive data integration project are identified (QMAS Report, 1997):

- The objective of data integration should be defined clearly from the start
- Data integration projects require a significant time commitment
- Barriers to participation must be identified and addressed
- Early financial commitment is a key to ensuring ongoing political commitment
- MIS (management information systems) staff should be involved from the start

In addition, there are some important questions regarding the use of multiple data sources from external organizations (QMAS Report, 1997):

- Are the data current enough to be useful?
- What are the content limitations of the data?
- What are the limitations in terms of available methodologies for analyzing the data?
- What are the technological requirements? What confidentiality issues are relevant?

All these questions should be carefully addressed in the data integration across multiple organizations.

It is worth noting that several cases in this paper are in health care field, it probably indicates that health care is a leader in data integration efforts. It also seems to us that organization-wide data integration is done for operational reasons, while data integration across multiple organizations (at least in the cases) is done for research and evaluation purposes.

References

Advanced Forest Technologies, Canada, Data Management with Integration of Multiple Data Source, 1997, <http://www.aft.pfc.forestry.ca/Proposal/dataman.html>.

Alciati, M. H., and K. Glanz, Using Data to Plan Public Health Programs: Experience from State Cancer Prevention and Control Programs, *Public Health Reports*, Mar/Apr96, Vol. 111, No. 2, 1996, pp.165-172.

Approaches to Data Collection, Chapter 5, 1997, http://nccic.org/research/nrc_dir/chap5.html.

Boss, L. P., and L. Suarez, Uses of Data to Plan Cancer Prevention and Control Programs, *Public Health Reports*, Jul/Aug90, Vol. 105, No.4, 1990, pp.354-360.

Child Care Bureau: Research, Data and Systems, 1998,
<http://www.acf.dhhs.gov/programs/ccb/data/index.htm>.

Clark County Recorder' s Office: Business Process Reengineering Project, 1998,
<http://www.co.clark.nv.us/recorder/bpr.htm>.

Culhane, D. P., and S. Metraux, Where to from Here? A Policy Research Agenda Based on the Analysis of Administrative Data, *Understanding Homelessness: New Policy + Research Perspectives*, Fannie Mae Foundation, 1997, pp. 341-360.

Dawes, S. S., Interagency Information Sharing: Expected Benefits, Manageable Risks, *Journal of Policy Analysis and Management*, Vol. 15, No. 3, 1996, pp.377-394.

Dingwall, J. [1], How to Make Your Organizational IQ Pay Off, *Canadian Government Executive*, Vol. 4, No. 2, 1998, pp. 20-22.

Dingwall, J. [2], Knowledge Management Approach for the Public Sector, *Canadian Government Executive*, Vol. 4, No. 1, 1998, pp. 14-17.

Drucker, P. F., The Next Information Revolution, *Forbes Magazine*, August 24, 1998,
<http://www.forbes.com/asap/98/0824/046.htm>.

Fuhs, F. P., Coupling Enterprise Planning with the Creation of the Conceptual Schema in Database Design, 1997, <http://hsb.baylor.edu/ramsower/acis/papers/fuhs.htm>.

Goodhue, D. L., M. D. Wybo, and L. J. Kirsch, The Impact of Data Integration on the Costs and Benefits of Information Systems, *MIS Quarterly*, September 1992, pp. 293-311.

Heimbigner, D., and D. McLeod, A Federated Architecture for Information Management, *ACM Transactions on Office Information Systems*, Vol. 3, No. 3, July 1985, pp. 253-278.

Huber, G., Organizational Information Systems: Determinants of Their Performance and Behavior, *Management Science*, Vol. 28, No. 2, February 1982, pp. 138-155.

Kentucky Kids Count, 1996, <http://www.louisville.edu/cbpa/kpr/kidscount/define96.htm>.

Lillquist, P. P., M. H. Alciati, M. S. Baptiste, P. C. Nasca, J. F. Kerner, and C. Mettlin, Cancer Control Planning and Establishment of Priorities for Intervention by a State Health Department, *Public Health Reports*, Nov/Dec94, Vol. 109, No. 6, 1994, pp.791-103.

Litwin, W., L. Mark, and N. Roussopoulos, Interoperability of Multiple Autonomous Databases, *ACM Computing Surveys*, Vol. 22, No. 3, September 1990, pp. 267-293.

Martin, J., Information Engineering, Savant Research Studies, Carnforth, Lancashire, England, 1986.

McKenna, D., Notes from the field, 1996,

<http://www.govtech.net/1996/gt/june/notesjune/notesjune.htm>

Muralidhar, S., The Role of Multiple Data Sources in Interpretive Science Education Research, *International Journal of Scientific Education*, Vol. 15, no. 4, 1993, pp.445-455.

O'Connell, J. J., et al., Health Care Data Governing Board: Annual Report, 1995, <http://www.ink.org/public/hcdgb/khcd95report.html>.

Planning Manual, Institutional Research & Planning, 1998, http://www.appstate.edu/www_docs/depart/irp/planning/appx-d.html.

QMAS workshop Report: Assessing Hospital Performance, written and produced by Severyn Healthcare Consulting & Publishing, Fairfax, Virginia, 1997, <http://www.qmas.org/guide-assessing/33total.htm>.

Robertson, P., Integrating Legacy Systems with Modern Corporate Application, *Communications of the ACM*, Vol. 40, No. 5, May 1997, pp.39-46.

SEI' s MassCHIP Program, <http://commonwealth2.Cam-colo.bbnplanet.com/dpd/dphhome.htm>.

Sheth, A. P., and J. A. Larson, Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, *ACM Computing Surveys*, Vol. 22, No. 3, September 1990, pp. 184-236.

Subrahmanian, V. S., S. Adali, A. Brink, J. J. Lu, A. Rajput, T. J. Rogers, R. Ross, C. Ward, HERMES: A Heterogeneous Reasoning and Mediator System, 1996, <http://www.cs.umd.edu/projects/hermes/overview/paper/section1.html>.

Van Den Hoven, J., Data Mart: Plan Big, Build Small, *Information Systems Management*, Winter98, Vol. 15, No. 1, 1998, pp.71-73.

**Center for Technology in Government
University at Albany / SUNY
1535 Western Avenue
Albany, NY 12203
Phone: 518-442-3892
Fax: 518-442-3886
info@ctg.albany.edu
www.ctg.albany.edu**