Overview of Research and Practices in Information Sharing for Enterprise Resource Planning

Jihong Zeng
New York Institute of Technology

Chengxiu Tang
Peking University

In recent years, the use of multiple data sources for enterprise resource planning and decision making have 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 resource planning and decision making. It identifies current research and practical experiences in the use of multiple data sources to support performance measurement, strategic planning, and inter-organizational business processes. 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 also 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.

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).

Knowledge and information-sharing networks are emerging in an increasing number of government programs and policy arenas. These inter-organizational and intergovernmental networks facilitate cross-program and cross-functional coordination and support communities of practice. They often include shared repositories of detailed program or administrative information, accessible to all participants that can address such needs as program evaluation, reference services, or technical assistance (Eglene et al., 2007). In today's global economy, enterprises need to better use their information resources to operate more efficiently and effectively. A study from the University of California at Berkeley found the amount of data organizations collect and store in enterprise database doubles every year, and slightly more than half of this data will consist of reference information, which is the kind of information strategic business application and decision support system demand (Kestelyn, 2002). 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). 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 inter-organizational 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 why we need to integrate multiple data sources and what are the benefits of integrating multiple data sources for enterprise-level planning and decision making. It also identifies and compares the benefits, barriers, and lessons learned that involve integrating different data sources 1) within one organization, and 2) across multiple organizations.

The Importance of Multiple Data Source Integration

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).

Examples of benefits of data integration include the cases of the 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 (Subrahmanian, et al., 1996). In each above 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. In this 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.

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).

EFFORTS INVOLVING ORGANIZATION-WIDE DATA INTEGRATION

In this study, a number of cases of efforts involving intra-organization data integration were identified, examined, and analyzed. We were especially interested and examined a few among them such as the legacy system integration in Kaiser Permanente (Robertson, 1997); tracking production schedules in Devlin Electronics; the cross-product line analysis in Southern Cross; the single customer interface created in Greenfields Products; and the better dispactching and shipment tracking system in Burton Trucking Company (Goodhue et al., 1992), etc.

For example, in the case of legacy system integration in Kaiser Permanente (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. 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.

In the case of Kaiser Permanente, 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 other opportunities for using legacy data, however, the pressure to take advantage of legacy data is extremely high.

Another case in organization-wide data integration is the tracking of production schedules 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 (Goodhue, et al., 1992).

Benefits of Organization-wide Data Integration

Based on the cases we examined and analyzed, it is found that organization-wide data integration tends to lead to the following major benefits in the context of enterprise-level planning and decision making: (1) improved managerial information for organization-wide communication; (2) improved operational coordination across sub-units or divisions of an organization; and (3) 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. It can have a positive impact on reducing costs by reducing redundant design efforts. 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 of Organization-wide Data Integration

Our case studies show that data integration can have a positive impact on reducing costs by reducing redundant design efforts. However, we found that 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 include: (1) compromises in meeting local information needs; (2) bureaucratic delays that reduce local flexibility; and (3) 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 on Organization-wide Data Integration

The following lessons have been derived from the cases of organization-wide data integration: (1) 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 (Robertson, 1997). (2) 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). (3) A single logical design for use across multiple sub-units can be difficult. The more subunits involved and the more heterogeneous their needs, the more difficult it will be to develop a single design to meet all needs. (4) Data integration may change the organizational information flows, and affect individual roles and organizational structure (Sheth and Larson, 1990). 5) The cost of designing and implementing data integration must be also considered, because it might be much higher than expected.

EFFORTS INVOLVING DATA INTEGRATION ACROSS MULTIPLE ORGANIZATIONS

In this study, we also identified and examined a number of cases on efforts involving inter-organization data integration. Some of the cases were from organizations in the same sector of the economy, for example, the cases of Five States' Cancer Prevention and Control Planning Model (Alciati and Glanz, 1996), and another Statewide Cancer Control Plan in Seven State's Health Departments (Boss and Suarez, 1990); other cases were from organizations in multiple sectors of the economy, for example, the cases of Quality Measurement Advisory Service in assessing hospital performance (QMAS, 1997), and the Health Care Governing Board in Kansas (O'Connell, et. al., 1995), etc.

In the case of Five States' Cancer Prevention and Control Planning Model (Alciati and Glanz, 1996), the five states, namely, Georgia, Maryland, North Dakota, Vermont, and Washington State, experienced using and integrating available data to develop cancer control plans for their states. While using health data to guide public health planning efforts was not new, information on how states use existing multiple data sources for comprehensive cancer prevention and control planning was limited. What was lacking was a clear picture of how these components fit together in a comprehensive state-level planning process, and how data are used to establish cancer prevention and control priorities and to identify proven interventions for implementation. Each state in this program used three categories of data: 1) Health data such as the number of state deaths, age-adjusted death rates for the state, and survival. 2) Behavioral data, including health behavior, risk factors, and determinants of behavior such as knowledge, attitudes, and beliefs. 3) 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, etc. For each type of data, the specific data source, the measures used, how the data were used to establish planning priorities and identify interventions were examined and analyzed. In these five state programs, comprehensive cancer planning efforts used a full range of integrated data, and linked these data to decision-making for cancer control. This case also provides a framework for public health planners to identify the type of data likely to be available for cancer prevention and control at the state level, and how they can be linked to public health planning (Alciati and Glanz, 1996).

Another example is the case of the Health Care Governing Board in Kansas (O'Connell, et al., 1995). 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. In 1995, the Governing boards implemented its first data collection initiative, a health system inventory. 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. It is not only an integrated data system of health care providers; it will also allow users to analyze data across professions and facilities from multiple data sources. The data is 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.

Benefits of Inter-Organization Data Integration

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; Health Care Governing Board in Kansas): (1) increased customers service quality; (2) increased existing personnel efficiency; (3) improved information quality, timeliness, and utilization; (4) increased accessibility, and analysis of information; and (5) the elimination of redundant data and tasks.

Barriers of Inter-Organization Data Integration

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. As summarized by Culhane and Metraux (1998), (1) legal restrictions may prevent access to a particular data set. (2) It is difficult to obtain 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". (3) Information sharing and data integration also need to deal with technical barriers. 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. (4) Integrating data systems frequently requires the concurrence of system administrators, directors of programs, and services consumers.

In addition, other barriers such as cost, timing, data standard and information overload also present challenges (QMAS Report, 1997). 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. It takes much more time to collect needed data from different sources across organizations. This time lag may cause synchronization problems. Different data standards will result in no clear vision of data strategy, and make the information decision support much more complex. 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. In spite of this enormous growth in enterprise databases, research from IBM revealed that organizations use less than 1 percent of their data for analysis (Brown, 2002).

Lessons on Inter-Organization Data Integration

Based on the experiences in the cases where organizations are either in the same or multiple sectors of the economy, some important lessons regarding the implementation of a

comprehensive data integration project are identified (QMAS Report, 1997). (1) The objective of data integration should be defined clearly from the start. (2) Data integration projects require a significant time commitment. (3) Barriers to participation must be identified and addressed. (4) Early financial commitment is a key to ensuring ongoing political commitment. (5) MIS (management information systems) staff should be involved from the start.

CONCLUSION AND FUTURE WORK

This study reviews uses of multiple data sources for enterprise resource 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 inter-organizational business processes. 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 also identified and compared.

It is worth noting that several cases mentioned in this study 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.

The purpose of this paper is to take the first steps towards the development of a methodology for integrating multiple data sources. The next step of our study is to examine and explore some important questions regarding the use of multiple data sources from external organizations, including 1) whether the data are current enough to be useful; 2) the content limitations of the data; 3) the limitations in terms of available methodologies for analyzing the data; 4) the technological requirements; and 5) what confidentiality issues are relevant. All these questions are hard to answer and they should be carefully addressed in the data integration across multiple organizations.

REFERENCES

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

Alciati, M. H. & Glanz, K. (1996). Using Data to Plan Public Health Programs: Experience from State Cancer Prevention and Control Programs. Public Health Reports, 111, (2), 165-172.

Boss, L. P. & Suarez, L. (1990). Uses of Data to Plan Cancer Prevention and Control Programs. Public Health Reports, 105, (4), 354-360.

Brown, E. (2002). Analyze This. Forbes, April 1, 169, (8), 96-98.

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

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

Dawes, S. S. (1996). Interagency Information Sharing: Expected Benefits, Manageable Risks. Journal of Policy Analysis and Management, 15, (3), 377-394.

Dingwall, J. [1] (1998). How to Make Your Organizational IQ Pay Off. Canadian Government Executive, 4, (2), 20-22.

Dingwall, J. [2] (1998). Knowledge Management Approach for the Public Sector. Canadian Government Executive, 4, (1), 14-17.

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

Eglene, O., Dawes, S. S. & Schneider, C. A. (2007). Authority and Leadership Patterns in Public Sector Knowledge Networks. The American Review of Public Administration, 37, (1), 91-113.

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

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

Heimbigner, D. & McLeod, D. (1985). A Federated Architecture for Information Management, ACM Transactions on Office Information Systems, 3, (3), 253-278.

Huber, G. (1982). Organizational Information Systems: Determinants of Their Performance and Behavior, Management Science, 28, (2), 138-155.

Kestelyn, J. (2002). No Longer an Afterthought. Intelligent Enterprise, 5, (13), 6.

Litwin, W., Mark, L. & Roussopoulos, N. (1990). Interoperability of Multiple Autonomous Databases. ACM Computing Surveys, 22, (3), 267-293.

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

McKenna, D. (1996). Notes from the Field. http://www.govtech.net/1996/gt/june /notesjune/notesjune.htm.

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

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

Robertson, P. (1997). Integrating Legacy Systems with Modern Corporate Application. Communications of the ACM, 40, (5), 39-46.

Sheth, A. P. & Larson, J. A. (1990). Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. ACM Computing Surveys, 22, (3), 184-236.

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

Van Den Hoven, J. (1998). Data Mart: Plan Big, Build Small. Information Systems Management, 15, (1), 71-73.