Center for Technology in Government

# Putting Information Together

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

## Abstract

# Putting Information Together

February 2000

The Center for Technology in Government (CTG), through the Using Information in Government (UIG) program, has worked with New York State agency project teams and partners from the public, private, and academic sectors to identify benefits and strategies for integrating and using information for program planning, evaluation, and decision making. The policy, management, and technology issues identified through our work with agency teams were shared with the public in a series of seminars focused on increasing the value of information to government programs. This report summarizes the presentations given at the fourth session of the Using Information in Government Seminar Series, "Putting Information Together: Building Integrated Data Repositories," which was held on February 9, 2000 at the University at Albany/SUNY.

# Introduction

The Center for Technology in Government (CTG), through the Using Information in Government (UIG) Program, has worked with New York State agency project teams and partners from the public, private, and academic sectors to identify benefits and strategies for integrating and using information for program planning, evaluation, and decision making. The policy, management, and technology issues identified through our work with the agency teams are being shared with the public in a series of seminars focused on increasing the value of information to government programs.

This report summarizes the presentations given at the fourth session of the **Using Information in Government Seminar Series**, **"Putting Information Together: Building Integrated Data Repositories,"** which was held on February 9, 2000 at the University at Albany/SUNY. This seminar dealt with the lessons learned during the Homeless Information Management System (HIMS) project conducted by CTG and the NYS Bureau of Shelter Services (BSS**)**. The prototype data repository incorporates information from a variety of sources to help BSS and homeless shelter providers manage and evaluate services for homeless families and single adults in New York State. Using the HIMS prototype development as an example, the seminar highlighted the management, policy, and technology issues organizations face when creating integrated data repositories. The seminar included several presentations, a demonstration, and panel discussion:

- BSS Director Robert Dawes and BSS Homeless Services Representative II Jane Wagner discussed how their agency developed the business case for the HIMS prototype.
- University at Albany Associate Professor Lakshmi Mohan talked about the data challenges of the project.
- Oracle Corporation Technical Manager Pat Schaffer recounted how the prototype data repository was built.
- The capabilities and functionality of the HIMS prototype were demonstrated by Oracle Corporation Principal
- Consultant Samir Ahuja with assistance from Bob and Pat.
- CTG Project Support Manager Donna Canestraro provided an overview of lessons learned during the HIMS project.
- The seminar concluded with a panel discussion moderated by Canestraro and involved Bob, Lakshmi, and Pat, as well as Greg Shinn (Director of Social Services for John Heuss House) and Edward Canfield (Data Processing Systems Auditor II for the NYS Department of Family Assistance).

# Key Points

A number of important concepts and issues were covered during the seminar.

- A vast majority, between 50 and 80 percent, of data warehouse projects fail because participants don't fully understand the value of integrated data and aren't sure how to use the combined information.
- Data repositories can integrate disparate sources of data to evaluate the impact of a set of programs and services on a specific population. Such repositories must resolve issues about: the intersection of programs and services, business drivers, data, technology, and the policy environment. They are built on a foundation of leadership, sponsorship, partnerships, and common vision.
- Data warehouse projects can be realized through the process of building a bridge from the original idea to the actual system. This is done by: defining a service objective that meets real business needs, developing a business case that shows how the project will work and what benefits it will provide, establishing partnerships with all key stakeholders, encouraging open communication among all participants, building and testing a data warehouse prototype of relevant data, resolving problems, and being persistent.
- Data by themselves are a worthless luxury unless they are used. Thus a prototype system, which converts data into actionable information, must be developed so that users can appreciate the value of the data.
- To define relevant data, focus on what an organization must do and know rather than what would be nice to do and know. The goal is to provide information that will be actively used for identifying and solving problems.
- The development of a data warehouse should employ an evolutionary approach in which each phase builds on the preceding one so that users derive value by using the data from each phase. Start with the core data elements and build a prototype system with real data, not dummy data, so that users can get a real feel for the information when they "test-drive" the prototype. If the prototype does not demonstrate value to the users, the project should be buried before more time and money are wasted on building a data warehouse that will not be used.
- Data marts have three components: data feeding, storage, and use. First, operations and external data are extracted, transformed, and fed into the warehouse. The data itself, and the meta data that describes the data, are then stored in the system. Finally, online analytical processing tools and applications are employed to use the data.
- Some of the challenges encountered when developing data marts include: extraction of data from multiple sources, data quality, ease-of-use, user support, data gaps, performance, implementation time frame, bad or missing data, and changing requirements.
- When doing data warehouse projects, 80 percent of the effort is in the data.
- Technology is important, but management and policy issues must also be addressed early in the development process. Organizations need to know who will be involved in the project, what stakeholder groups will be affected by it, what the system will do, where the data will come from, and how it will be obtained. They also need to create business rules, standard data definitions, confidentiality and privacy policies, and other guidelines for data use.

# Welcome

**Theresa Pardo - CTG**

Organizations spend millions of dollars building data warehouses, but as many as 50 to 80 percent of those projects fail. Most often they fizzle because planners and participants lack a clear understanding of why access to integrated information is valuable and how the combined information will be used. This astounding failure rate brings home the point that organizations must fully research and understand their business issues, policy environment, and information needs before attempting to build a sophisticated technology solution.

CTG's two-year Using Information in Government (UIG) Program recognizes that information is a valuable governmental asset and explores ways public sector agencies can best conduct information-intensive work. The UIG Program involves projects with seven New York state and local agencies to examine the policy, organizational, management, and financial factors that shape government's ability to get full value out of the

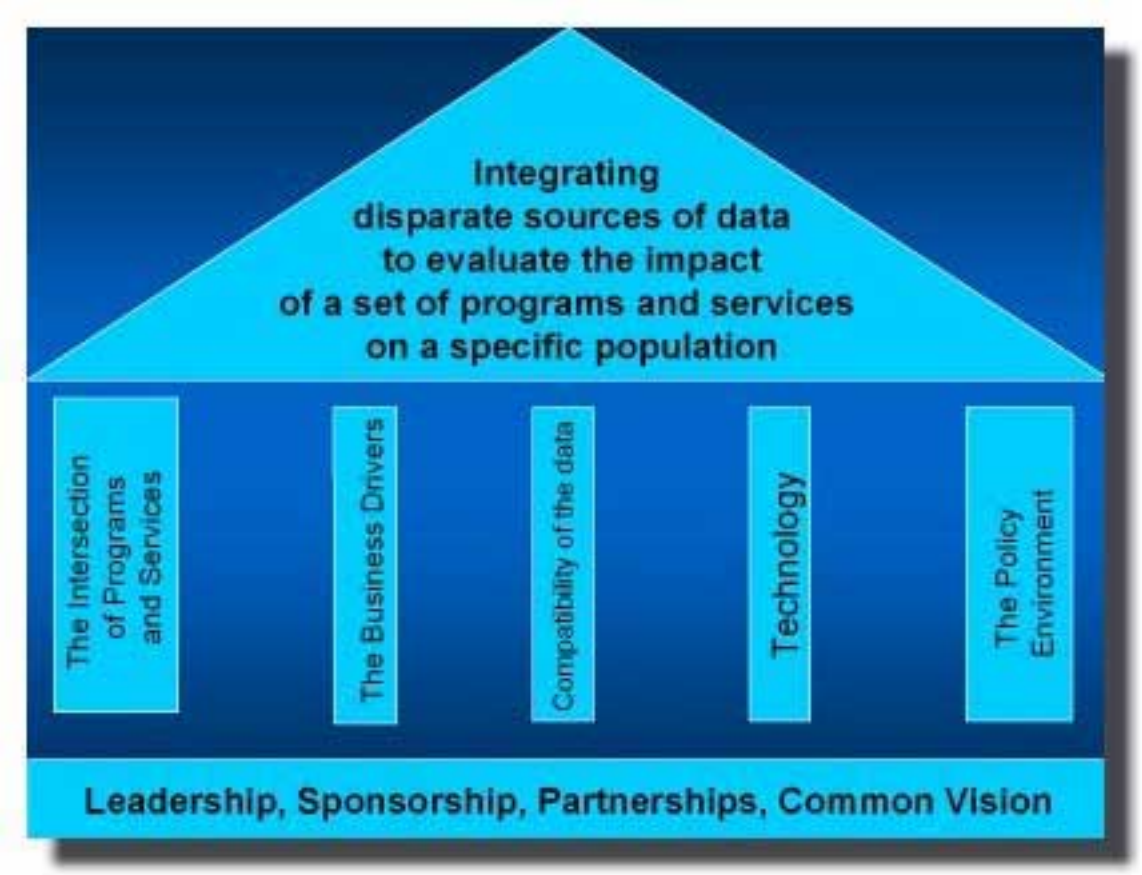
 

information it collects, creates, and maintains. Those seven New York agencies are:

- Office of the State Comptroller, Division of Municipal Affairs
- Central New York Psychiatric Center
- Office of Temporary and Disability Assistance, Bureau of Shelter Services
- Office of Real Property Services
- Office of the State Comptroller
- Department of Transportation
- New York City Department of Telecommunications and Information Technology

The results of these projects, all of which will be finished by June 2000, will be shared in the UIG Resource Kit-a Web- based collection of materials that will provide advice, experience, tools, techniques, and guidelines to help government managers become more effective information users.

By working with the NYS Bureau of Shelter Services on the Homeless Information Management System integrated data repository, CTG developed a model of the issues involved in this type of project. Data repositories usually integrate disparate sources of data to evaluate the impact of a set of programs and services on a specific population. The success of such repositories requires solving issues about: the intersection of programs and services, business drivers, data compatibility, technology choices, and the policy environment. Such systems are built on a foundation of leadership, sponsorship, partnerships, and common vision. The interorganizational nature of HIMS made these issues particularly critical.




The prototype process showed BSS and homeless service providers that a tool such as HIMS could be developed to provide them with the information they need to track homelessness and more effectively manage and evaluate the diverse services delivered to the homeless population. It also demonstrated that many critical factors must be addressed.

The prototyping process involved intense investigation into the policy, management, and technology barriers encountered when integrating multiple sources of data into a single repository designed to support system-wide assessments of services, as well as the changes required to overcome those barriers.

# Homeless Information Management System: Building the Business Case

**Robert Dawes - BSS**

**Jane Wagner - BSS**

Every night 30,000 New Yorkers go to bed homeless. Shelters approved by BSS temporarily house 20,000 people, 12,000 of whom are children. Homeless services have come a long way since the 1970s when shelters were primarily concerned with providing "3 hots and a cot"-three hot meals a day and a bed to sleep in at night. Today, homeless people are given a full range of services including preparation for permanent housing, education, health care, job training, counseling, and life skills. New York State spends $330 million a year tackling the problem of homelessness.

BSS needed a more effective way to manage programs, evaluate services, and devise strategies for helping people permanently break out of the cycle of homelessness. The agency's first step was to define a service objective: develop a uniform automated database to provide management with information needed to track homelessness and to more effectively manage the program.

Out of that objective came specific project goals:

- Collect demographic and service information about homeless people
- Evaluate effectiveness of services provided
- Use existing electronic data from several sources
- Make the system voluntary for providers
- Use results to do program planning at all levels by sharing information with providers and local districts

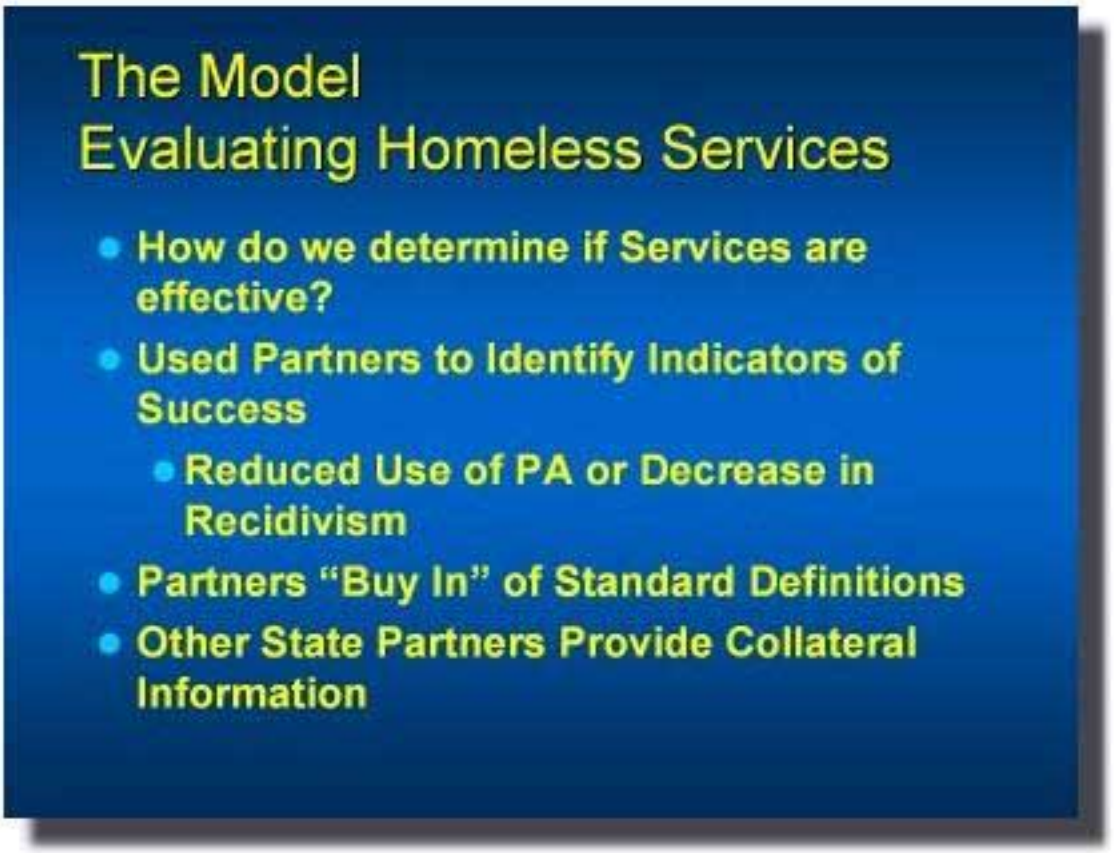
In order to meet the service objective and fulfill project goals, BSS had to engage homeless shelter providers and local social service districts. BSS was able to secure the support and participation of these groups through open communication and taking the project idea directly to the stakeholders. In addition to inviting the primary stakeholders to be full partners in the project, the team worked hard to involve shelter providers and local districts from the beginning. The team earned the trust and support of the stakeholders by assuring them that the system would not be mandatory, and demonstrating how it would benefit their work.

Once everyone was on board, the project team had to work through several issues regarding confidentiality and data. BSS worked closely with providers about:

- Protecting client information through specific agency protocols
- Restricting access to individual shelter data to the shelter itself
- Making aggregate data available to all users
- Establishing internal and external controls
- Including only demographic and service data in the system
- Excluding case notes from the system
- Creating key identifiers to standardize information

The team developed a model for how they would evaluate homeless services. This model involves using shelter providers and local social service districts to identify indicators of success. Those indicators include reduced use of public assistance or a decrease in recidivism in the shelter system. The model also relies on providers using the new standard definitions and other state agencies providing collateral information. These requirements went into the HIMS prototype, which focuses on a subset of the family shelter population data.




All of the relationships that were forged, issues that were resolved, and discussions that took place helped BSS build its business case for the HIMS system. That business case is being used to prove the value of an integrated system to track and manage homeless services. The case is also an important tool in the effort to secure support and funding to build the full HIMS system and implement it across the state.

The most important lesson learned by BSS involves the process of building a bridge from the original idea to the actual system. The team feels it did this by defining a service objective that meets real business needs, developing a business case that shows how the project will work and what benefits it will provide, establishing partnerships with all key stakeholders, encouraging open communication among all participants, building and testing a prototype, resolving problems, and being persistent.


# The Data Odyssey of HIMS

**Lakshmi Mohan - University at Albany**

Building a data warehouse is a complex process that is beset by problems. It takes persistence to see the project through to the end. And as with any data warehouse project, the HIMS project was a data odyssey.

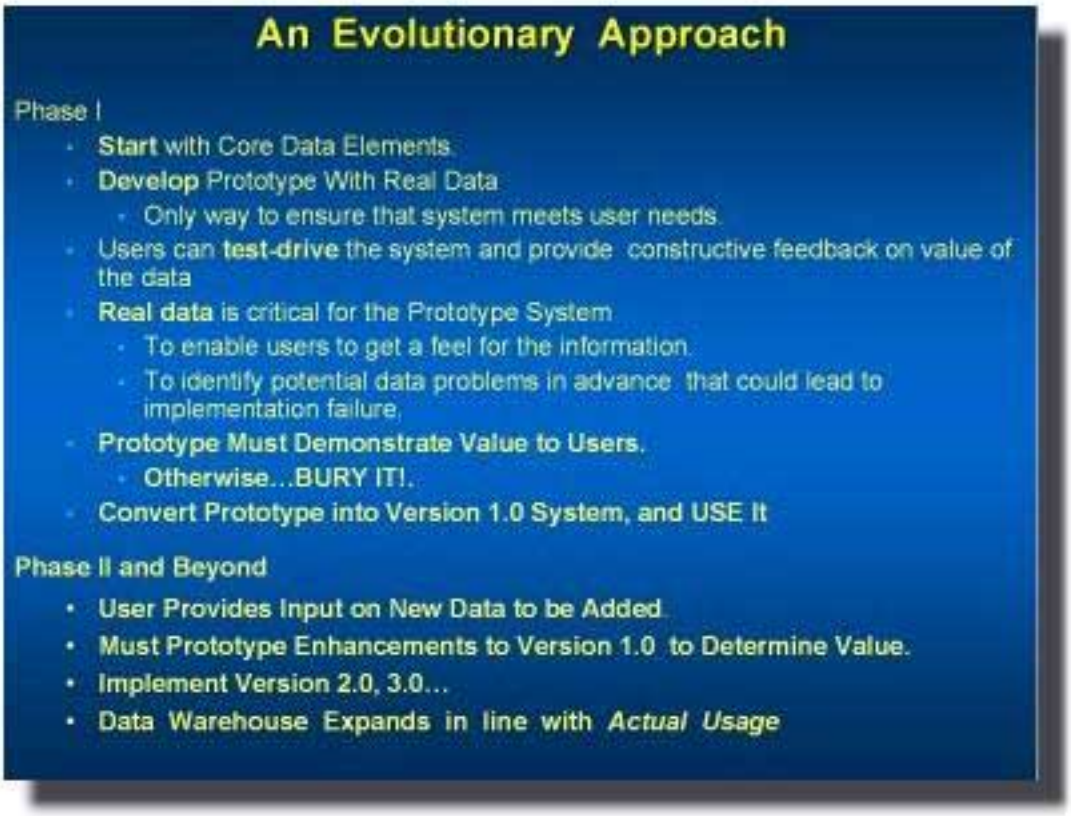
The objective of a data warehouse project is to create an integrated database of relevant data from multiple sources to support managerial decision-making. The challenge is to define relevant data, which must be done in the context of how the data will be used by management for problem-finding and problem-solving. The relevant data are determined by focusing on what you must do and know, rather than what is nice to do and know. Since you can't use every possible piece of data, you must identify the data that is necessary for your business needs. The key is to get actionable information.

Another powerful concept that must be applied in a data warehouse project is the "satisficing" concept. Better and more data will entail higher costs and more development time. However, the value of the project depends on the impact of the data on managerial actions. The satisficing approach to building data warehouses advocates that management only needs information that is "good enough" (satisficing) for decision making, because "perfect" information would cost too much and take too much time.


The process of actually building the data warehouse should employ an evolutionary approach. In this approach, the data warehouse evolves over time with users getting value in each phase because they actually use the data. To ensure that the data have value, a prototype system should be built in each phase of the project.




In the first phase, you must start with the core data elements that are critical for managerial decision-making, and develop the prototype using real data. The prototype is the only way to ensure that the system built on this data meets user needs. It gives users a real feel for the information produced by the system, allows them to give constructive feedback, and identifies potential data problems that could lead to implementation failure. The prototype must demonstrate the value of the data to users. If they do not see the value, then you should discard it before more time and money is wasted on building a data warehouse that will not be used. But if users believe that the system will help them do their jobs better, then you should move forward by converting the prototype into version 1.0 and install it in the users' workstations.

In the second phase, version 1.0 is already in full operation. Users can now provide input on new data to be added to the system that will enhance its utility. These enhancements to version 1.0 should also be prototyped to determine their added value. Implement version 2.0 and continue, if users desire it, with version 3.0. Continue this same process with version 3.0 and so on. The data warehouse should expand in time with its actual use. The key is to start small and keep going until you have a comprehensive data warehouse that meets business needs.

The data are the foundation of the HIMS prototype. The integration of data from several sources posed several challenges in this project.

- A comparative analysis of data elements from the files of two shelter providers revealed gaps in the records. BSS had to send staffers out to the providers to collect the missing data from paper files on clients. The team focused on gathering data about demographics and the following services: housing preparation, employment, child care, mental health, substance abuse, and living skills.
- Some of the information was inconsistent between the shelters. For instance, there were more than a dozen terms used to describe clients' ethnicity, and neither shelter used all the same vocabulary. Common data definitions had to be created to solve these kinds of issues.
- Relevant data from the state's Welfare Management System also had to be incorporated into the

prototype. This involved gaining a thorough understanding of the system's functions and features to determine which payment and benefits information were necessary to meet the business needs of the users.
- Key metrics were defined to help sort out the data. For example, the team devised definitions for "first time" and "repeater" clients. And the group created a formula to compute the length of stay for current residents.

Defining the business rules for handling a variety of data problems encountered in the real data was the key to integrating the data from the different sources. The business rules were defined on the basis of the system design that was created to convert the data into an interactive on-screen reporting system. The system design was critical for converting raw data in the data warehouse into actionable information that management could use. Without the system, the data warehouse would be a worthless luxury.

Once the database was defined and the system design was specified, the next step was to actually build the prototype.

# Building the HIMS Prototype

**Pat Schaffer - Oracle Corporation**

CTG corporate partner Oracle Corporation took the lead in building the HIMS prototype, which was designed to: collect data from different sources, place it in a single repository, clean and transform it, provide fast and easy-to-use views of it, and answer questions about homeless clients, services, and programs.

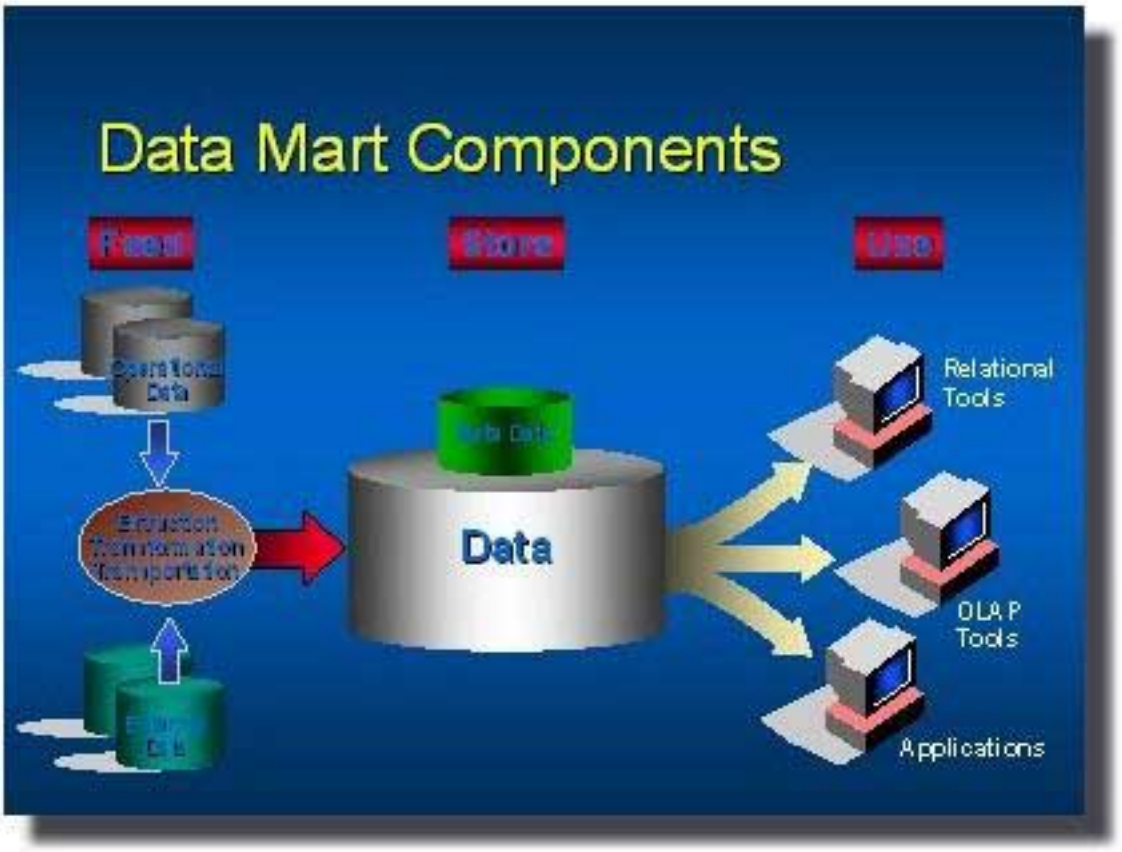
The prototype was built using Oracle data mart methodology, an iterative cycle of five development steps and 13 processes. Data marts are repositories of integrated, non-volatile, and time-variant data collected from heterogeneous transaction processing applications and stored in a format optimal for reporting and strategic analysis for a specific business area.


HIMS, like other data marts, uses online analytic processing-**or OLAP** - applications that report and summarize data for ad hoc reporting and decision support.

Data marts are systems that enable organizations to retrieve information to make crucial business decisions. The systems work on the feed-store-use model. Operations and external data are extracted, transformed, and transported into a storage facility; it is vital that these three processes be done separately. Selected data from operational systems are extracted and transported to the target database environment. The data are also transformed, meaning they're filtered and integrated into consistent and uniform formats in the target database. Relational and OLAP tools and desktop applications are then used to query the data to find specific information.




In the case of HIMS, a staging area houses 4,800 cases over three years from two shelter providers in New York City, as well as data from BSS and the state Welfare Management System. The users from BSS, shelter providers, and local social service districts will be able to access the system through the Internet. OLAP tools put the data in a three dimensional cube structure that provides different views of the information-by programs, time, facilities, clients, services, and providers. These tools provide: a multidimensional view of the data, drill-down analysis, analytical functions, access to detail and summary data, and integrated graphing and charting. This type of application allows all levels of system users to conduct queries and analysis.

Building a successful data mart presents a host of challenges that must be dealt with, such as data quality, user support, ease of use, performance, bad or missing data, implementation time frame, and changing requirements. But there's a list of data mart tips that will help you conquer these challenges:

- Simplify, simplify, simplify.
- A project never ends if it's successful.
- If it's not quality, they will not come.
- All end-users are not alike.
- Meta data is mega important.
- Encapsulate your extraction, transformation, and transportation.
- Capture data samples early on.
- The devil is in the detail.
- Beware of online transaction processing (OLTP) creep.
- There is method to this madness.

In addition, a number of helpful references about data marts exist. For example:

- "Data Warehouse Lifecycle Toolkit" by Ralph Kimball, 1998, ISBN 0-471-25547-5
- "Data Warehouse Toolkit" by Ralph Kimball, 1996, ISBN 0-471- 15337-0
- "Oracle 8 Date Warehousing" by Gary Dodge and Tim Gorman, 1998, ISBN 0-471-19952-4

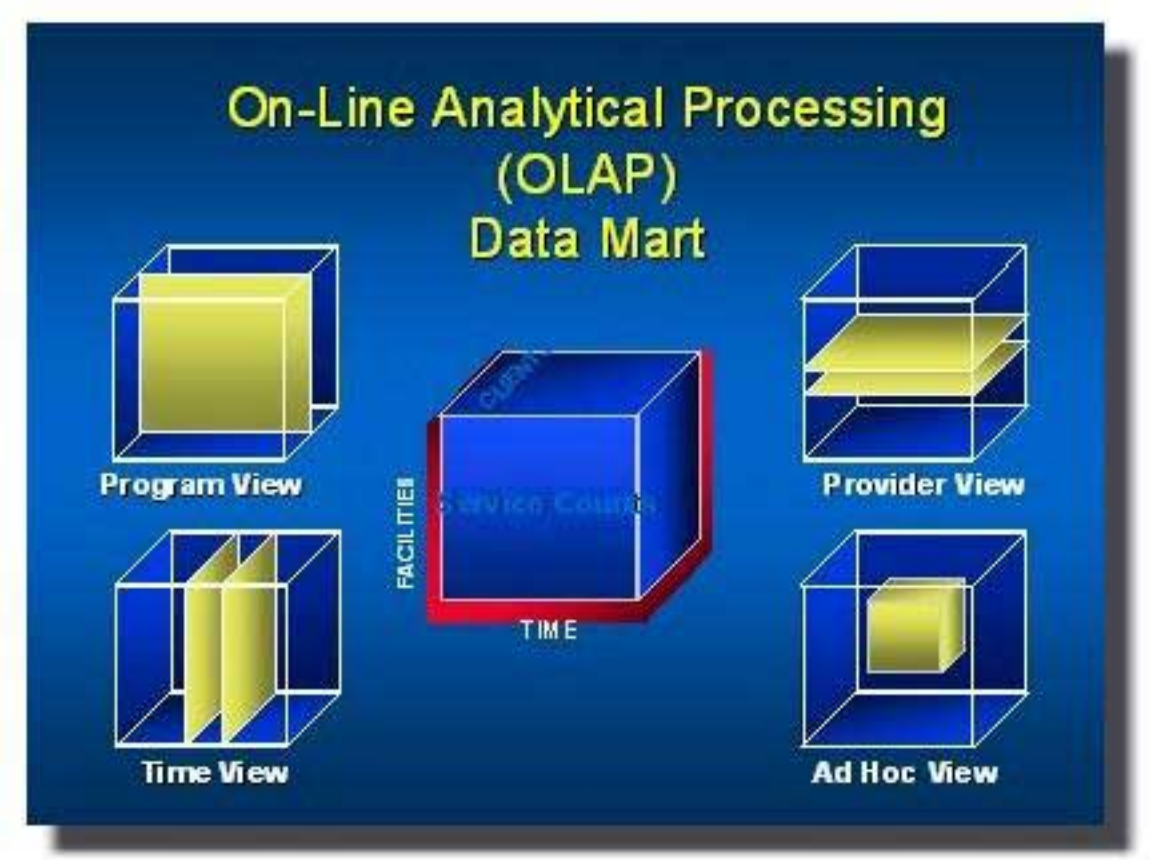

# HIMS Prototype Demonstration


**Robert Dawes - BSS**

**Pat Schaffer - Oracle**

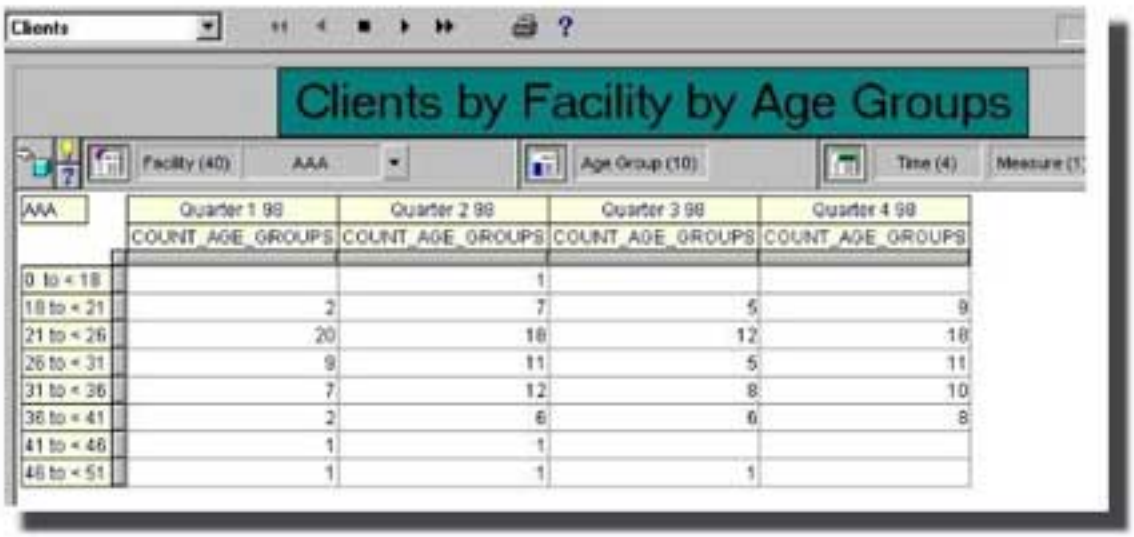
**Samir Ahuja - Oracle**


The HIMS prototype runs on **Oracle Express 6.0** and extracts data from an **Oracle 8** database. Data were taken from various sources, consolidated or aggregated based on business rules, and displayed in cubes. The prototype summarizes transactions into multidimensional views ahead of time. And because the data are summarized or aggregated at the time they are loaded into the system, user queries are very fast. These data cubes display the HIMS information in several views, including by program, provider, time, and ad hoc.



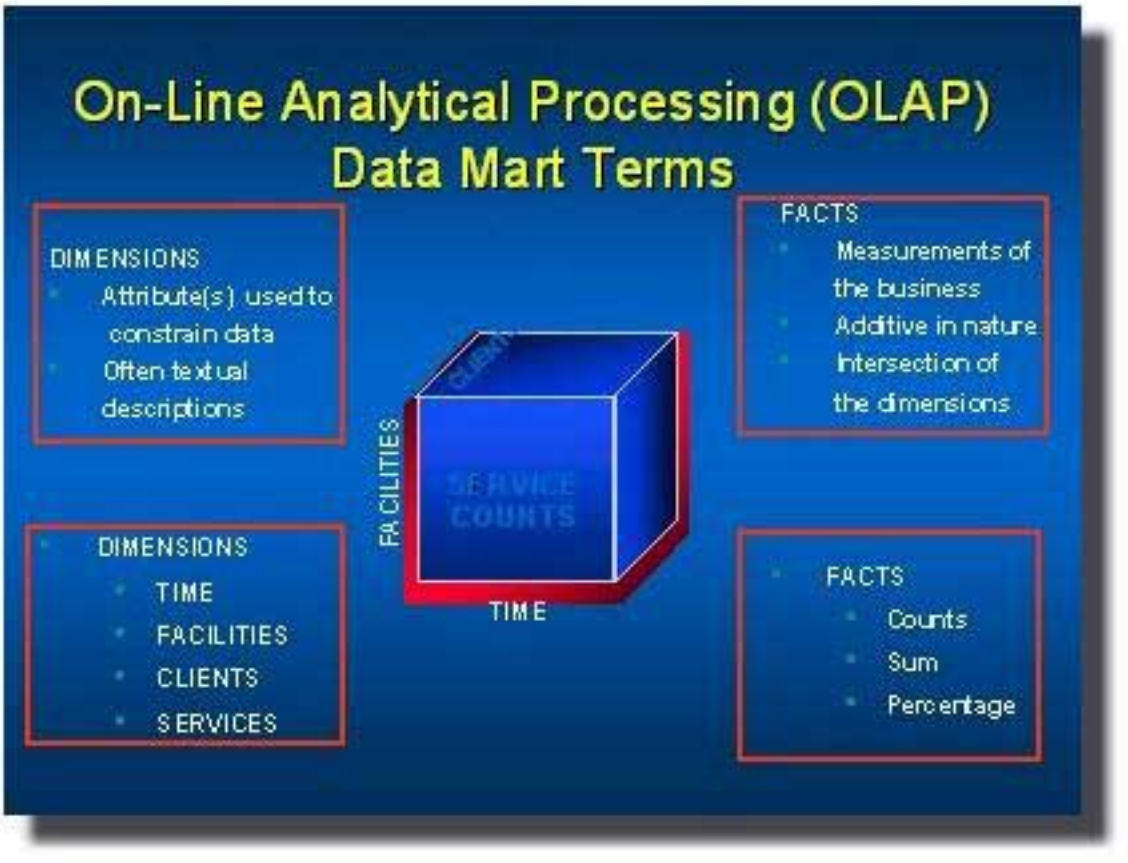

The prototype includes such user-friendly features as drop-down menus, spreadsheets, graphs, and drill-down capabilities. For example, users can view the client population, segmented into age groups, in a specific homeless shelter over the course of a year.




| AAA | Quarter 1 98 | Quarter 2 98 | Quarter 3 98 | Quarter 4 98 |
|---|---|---|---|---|
| | COUNT_AGE_GROUPS | COUNT_AGE_GROUPS | COUNT_AGE_GROUPS | COUNT_AGE_GROUPS |
| 0 to < 18 | | 1 | | |
| 18 to < 21 | 2 | 7 | 5 | 9 |
| 21 to < 26 | 20 | 18 | 12 | 18 |
| 26 to < 31 | 9 | 11 | 5 | 11 |
| 31 to < 36 | 7 | 12 | 8 | 10 |
| 36 to < 41 | 2 | 6 | 6 | 8 |
| 41 to < 46 | 1 | 1 | | |
| 46 to < 51 | 1 | 1 | 1 | |

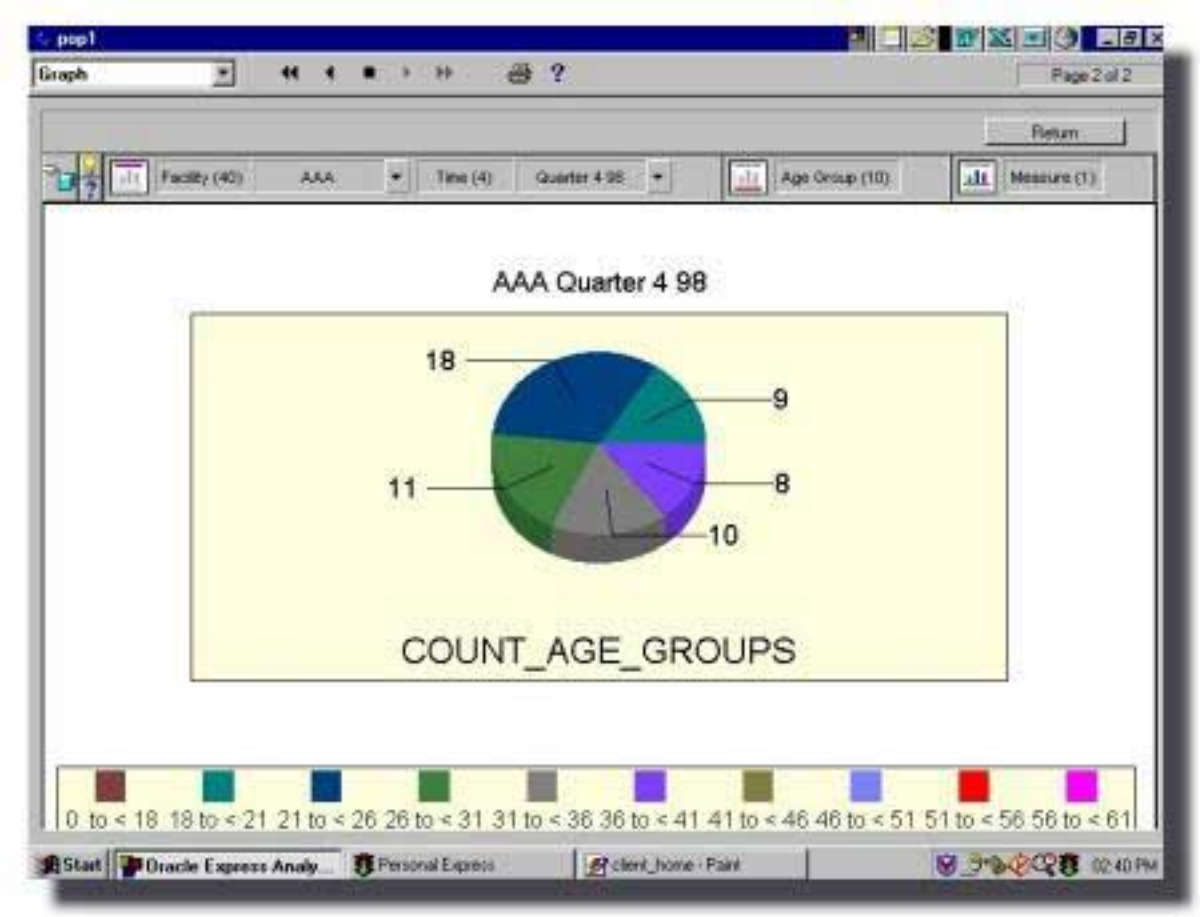
The ad hoc queries enable users to access specific slices of information. Data are available from the aggregate level all the way down to specific client populations and individual shelters. Users can search based on age, education level, family size, individual shelter, provider, and so on.




The versatility of the prototype system, which also produces graphs, allows users to access a wealth of

information for analysis and decision making. The power of the HIMS data mart lies in allowing users to view the data according to their unique perspective and needs.




# Lessons Learned

**Donna Canestraro - CTG**

Looking back on the project, the HIMS team articulated a number of the lessons learned during the prototype development process. Some of these lessons may prove useful for other organizations that are considering building their own data repositories.

- More data issues arise as more data sources are utilized.
- Business rules need to be developed throughout the process.
- Data definitions and standards need to be decided throughout the process.
- Collaboration among data providers and the state will continually need to be fostered.
- The system can be developed to answer a core set of questions proving value to the state and provider community.
- Communication channels need to be identified and maintained throughout the process.
- Avoiding negative key stakeholders will cause the project to fail.
- There is more to the cost of the system than just the hardware and software.
- It's important to recognize all cost categories to ensure the success of the project.
- All issues and barriers need to be reviewed from policy, management, and technology perspectives. Each should be revisited as the project progresses.
- People who understand the project and the data need to be part of the project team.

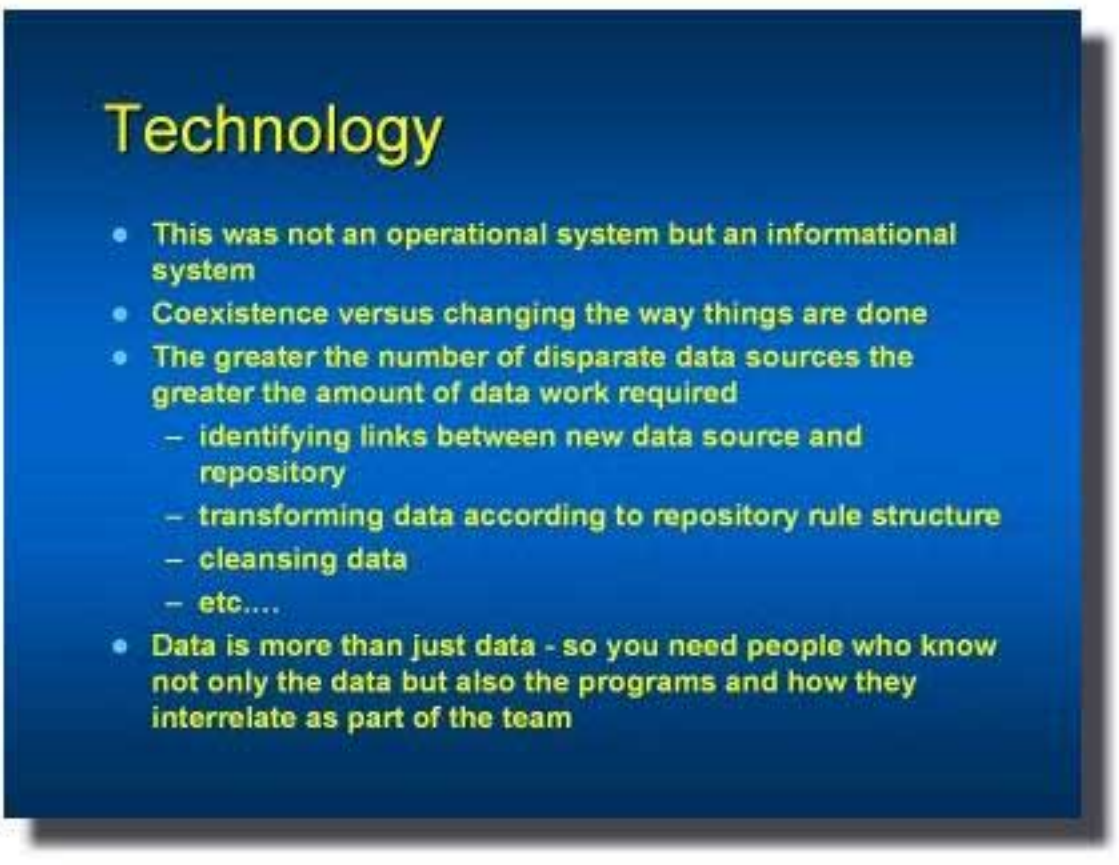
- The project team should be comprised of people with the authority to make decisions about the project progress, as well as any necessary program changes.
- Acknowledge that the infrastructure is key at all levels, including hardware and software procurement, as well as training, administrative, and development costs.

The team learned these lessons when dealing with the management, policy, and technology issues surrounding the project.

The management issues focused on the "who, what, where, and how" of the project. Numerous discussions considered: who should be involved, who the stakeholders are, what roles need to be filled, what the system will do, where the data will come from, and how to obtain it. It's vital to answer these questions in the early stages of the project.

Policy issues also garnered a lot of attention. Business rules were agreed upon. Data and their definitions were standardized. Confidentiality policies were applied to protect clients. Multiorganizational data sources were included to ensure a robust data sample. And a service evaluation model was developed collaboratively by BSS and a group of service providers with assistance from CTG.

After identifying the management and policy issues, the team was prepared to tackle the technology end of the project. This involved agreeing that the prototype is an informational system for tracking, trending, and analyzing homeless services. The team realized that the amount of data work would increase as the number of disparate data sources increased. This work involved identifying links between new data sources and the repository, transforming data according to the established rule structure, and cleansing the data. Project participants also recognized that you need people on the team who understand the data, as well as the programs and services that they describe.




By answering the key management, policy, and technology questions, the team was able to effectively

design and build the HIMS prototype. Some of the critical factors leading to the success of the project were:

- Participants continuously invested time in building relationships to ensure the engagement of the necessary local, nonprofit, and state partners throughout the process and beyond.
- The prototype affirmed the idea that potential users saw the system as valuable and therefore would invest in its success.
- The team members possessed the right mix of skills necessary to make key decisions and resolve issues quickly and effectively.
- Participants invested time at the beginning of the project to fully understand the key questions: what was to be achieved and why, who needed to be involved and why, and what were the best ways to achieve the goal and why.
- The team clearly defined the purpose and scope of the prototype project.
- Strong project management techniques were utilized at several levels-internally with the project team and externally at homeless provider meetings.
- Participants matched the technology to the job.
- The team followed a prescribed methodology in the process and technological areas.

# Panel Discussion

**Moderator:** Donna Canestraro

**Panelists:** Robert Dawes, Lakshmi Mohan, Pat Schaffer, Greg Shinn, Edward Canfield

A series of questions posed by the moderator provided the framework for a lively discussion of the challenges of integrating information to build data repositories.

Greg is an active member of the local provider community in New York City. He discussed the factors that were critical to his continued participation in the HIMS project:

- It takes vision. The organization has to position itself to further its own mission. In his shelter's case, leaders realized they needed to be at the forefront of technology in order to best serve their clients. They started this effort by establishing their own database to track services and clients.
- When he heard about HIMS, he realized that getting in on the ground floor of the project would give his organization an opportunity to influence the end product. Participation in HIMS enabled him to make sure the system would include his shelter's information and meet his business needs.

Ed has a deep knowledge of the NYS Welfare Management System and was a critical member of the project team. He commented on key data issues:

- Data queries have to come into line with the project. It's vital to understand where the data comes from.
- Meta data is also very important. You have to know the data and what they mean.

Lakshmi has worked with many organizations in similar efforts to use information in new ways to support decision making and planning. She compared the HIMS project to others she has worked on in the public and private sectors:

- She compared a data warehouse project for Kaplan Education Centers to her work on HIMS. Both involved the same kind of work, especially data issues.
- Cleansing, defining, and standardizing data are involved in every data warehouse project. Public or private, large or small, all organizations have to deal with these issues.
- You have to look at the actual data and understand it before you can resolve these issues and design a good system.

# Presenter and Panelist Contact Information

**Samir Ahuja**

Principal Consultant, Oracle Corporation 7 Southwoods Boulevard, Albany, NY 12211

Phone: 518-257-7408 Fax: 518-432-6841

Web site: www.oracle.com

**Donna Canestraro**

Project Support Manager, Center for Technology in Government 1535 Western Avenue, Albany, NY 12203

Phone: 518-442-3892 Fax: 518-442-3886

Web site: www.ctg.albany.edu

**Edward Canfield**

Data Processing Systems Auditor II, Bureau of Audit and Quality Control, NYS Department of Family Assistance 40 North Pearl Street, Riverview Center, Albany, NY 12243

Phone: 518-402-0027 Fax: 518-402-0061

Web site: www.otda.state.ny.us

**Robert Dawes**

Director, Bureau of Shelter Services, NYS Office of Temporary and Disability Assistance 40 North Pearl Street, Floor 8B, Albany, NY 12243

Phone: 518-486-7732 Fax: 518-474-7058

Web site: www.otda.state.ny.us

**Lakshmi Mohan**

Associate Professor, School of Business, University at Albany/SUNY 1400 Washington Avenue, Albany, NY 12222

Phone: 518-442-4927 Fax: 518-442-2568

Web site: www.albany.edu

**Theresa Pardo**

Project Director, Center for Technology in Government 1535 Western Avenue, Albany, NY 12203

Phone: 518-442-3892 Fax: 518-442-3886

Web site: www.ctg.albany.edu

**Pat Schaffer**

Technical Manager, Oracle Corporation 7 Southwoods Boulevard, Albany, NY 12211

Phone: 518-257-7408 Fax: 518-432-6841

Web site: www.oracle.com

**Greg Shinn**

Director of Social Services, John Heuss House 42 Beaver Street, New York, NY 10004

Phone: 212-785-0770 Fax: 212-785-0776

**Jane Wagner**

Homeless Services Representative II, Bureau of Shelter Services, NYS Office of Temporary and Disability Assistance 40 North Pearl Street, Floor 8B, Albany, NY 12243 Phone: 518-402-3599 Fax: 518-474-7058

Web site: www.otda.state.ny.us