RESEARCH REPORT

# DATA WAREHOUSING, FLOW MODELS, AND PUBLIC POLICY

Prepared by Erin Dalton, Wilpen Gorr, Jennifer Lucas, John Pierce

# Contents

# Abstract

The Department of Human Services (DHS) of Allegheny County, Pennsylvania is perhaps the first such organization in the country to build and use a data warehouse—a system that integrates data from several sources and makes it available for decision-making.

The result of consolidation in the mid-1990s, DHS has five program offices with a budget over $750 million and oversees approximately 600 service providers. Policy and management activities include the usual gamut of budgeting, facility location, cost/benefit analysis, and so forth—and necessarily must cut across departments. DHS has been a pioneer in the cross-system movement in human services delivery, drawing on resources from several departments internally and externally (70 percent of clients obtain services from more than one department).

Decision-making needed to be objective and databased where possible; however, departmental operational computer systems were incompatible and could not be integrated to provide cross-system data. DHS' Data Warehouse is the solution for integrated data.

This paper reviews the context of decision-making at DHS; provides a brief overview of data warehouse technology; reviews the DHS Data Warehouse data holdings and unique public sector data issues; provides several data warehouse applications to policy and management decision-making; and concludes with future work. Clearly, data warehousing is an ideal technology for use in many local government settings, enabling objective and comprehensive decision-making.

# Introduction

The Allegheny County Department of Human Services (DHS) is one of the first such organizations in the country to have a data warehouse—a technology that integrates data from several sources and makes it available for decision making. Funded initially by local grant-making foundations, this innovation was critical for policy making and management after the county's human service programs were joined under the umbrella of a single department.

The purpose of this paper is to describe the context and need for integrated data in human services, describe data warehousing technology and its unique challenges in the public sector, describe some innovative data warehouse applications that have arisen in part from joint projects with universities, and discuss future work. Data warehousing is a promising information technology in support of public sector policy and decision-making, and it promises to directly support service delivery transactions in the future.

There are 3,143 counties or county equivalents in the United States, most of which are the primary providers of human services and account for the majority of county government expenditures. These services provide a floor on endowments for the less fortunate members of society. For example, DHS—in the 28th largest county with a population of 1.28 million—has programs providing services in the following categories:

• Aging,
• Mental health,
• Drug and alcohol,
• Child protective,
• At-risk child development and education,
• Hunger,
• Emergency shelters and housing for the homeless,
• Energy assistance,
• Non-emergency medical transportation,
• Job training and placement for youth and adults, and
• Individuals with mental retardation and developmental disabilities.

In the mid-1990s, DHS was formed as a result of a major consolidation of human services departments in Allegheny County, with the original existing now under the umbrella of DHS. Figure 1 illustrates the organization of DHS' five major program offices, which oversee service deliveries from more than 600 county and nonprofit service providers. In 2004, DHS served 231,400 persons, employed 1,052 staff members, and operated with a budget of $757 million from 80 funding sources.

## Introduction

Consolidation, coupled with the Data Warehouse, has facilitated a cross-system movement in human services delivery in Allegheny County. Efforts, like Community Connections for Families (CCF) and Family Group Decision-Making, draw on a wide range of services from within and outside of DHS to address the needs of family units. Prior to such programs, county human services focused on individuals within single agencies, which lead to much duplication, inefficiency, and ineffectiveness.

The Data Warehouse was instrumental in the effective reconfiguration of services and has been the basis for several joint research projects with Pittsburgh-area universities and research organizations (e.g., RAND), through funded projects and student projects. These projects are contributing to policy-level knowledge on the behavior of targeted populations in response to existing or planned programs.
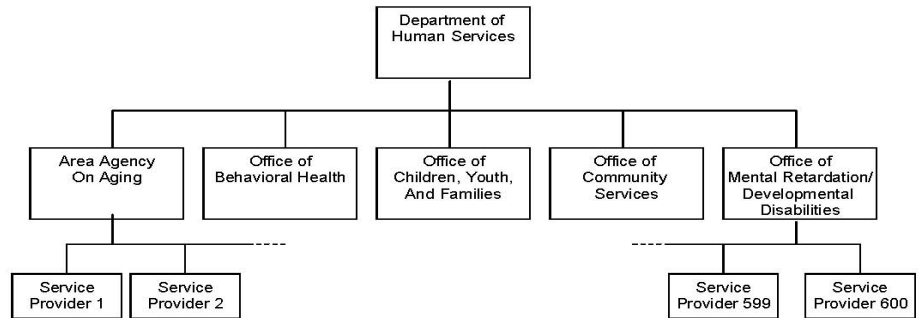
**Figure 1: Composition of Allegheny County Department of Human Services, post-1997**

# Data Warehousing Technology

Data warehousing is a technology dating back to the late 1980s and early 1990s that was developed in response to both the growing need for management information and the inability of operational information systems to deliver it (Data Warehouse, 2006). Operational systems maintain records on service delivery transactions, clients, employees, etc. within separate units. Enterprises generally have many operational systems that are isolated and non-integrated. As a result, decision making was not well supported by data across operating units. Also, operating systems were not configured for compiling or reporting management-level information. Data warehousing went through several phases in response to such needs. Today the typical data warehouse, including the DHS Data Warehouse, is offline and gets periodic data updates (daily, weekly, or monthly) from operational systems. Already available today, and a next step for DHS, is a real-time data warehouse that obtains data as soon as it is entered in operational systems.

**Definition**
There are several definitions for a data warehouse. W. Inmon, a founder of data warehousing, described it as follows (1995):

*"A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process."*

- <u>Subject-oriented</u> means that all data elements or attributes on a subject, such as service deliveries or clients, are collected from various sources. It also means that records are extracted on subjects from transactions and related records.
- <u>Integrated</u> means that data are collected from many sources and merged into a coherent whole.
- <u>Time-variant</u> refers to data collected on subjects over time, such as service deliveries to the same client over time.
- <u>Non-volatile</u> means that data is not erased over time. Instead, new records are added.

As is the case with the DHS Data Warehouse, the collection of data is the one common and consistent source of information about the entire enterprise and its affiliated units. Its data have been integrated and processed to eliminate duplicates, cleaned of errors and inconsistencies, made consistent over time, and so forth. It's the single information source that ensures that every member of the enterprise is "on the same page." It's also the only place for information that integrates across all units; for example, on all services that a client receives from DHS.

# Data Warehousing Technology

**Components**

Figure 2 illustrates the components of a typical data warehouse. The sources of data are the operational systems maintained by the operational units of an enterprise. In government, computers, operating systems, and software typically vary from unit to unit, making the next step — data transformation — more challenging. Data warehousing packages, such as the Cognos package used by DHS, have extensive import capacity to input and transform data from many sources into a common format. Besides changes in data format, it is at this stage that a series of steps are used to clean, standardize, and otherwise prepare data for storage in the data warehouse.

There are two schools of thought for the database schema, or table design, for data warehouses. One is to compile and store data aggregated to "data cubes" which contain counts, sums, averages, etc. by dimensions of a hypercube; for example, by time period, product/service, client type, and location (Kimball and Ross, 2002). The other school of thought is to store the imported and cleaned data in normalized data form and at the individual transaction level, in related tables, which are the input to aggregation. This is called an "operational data store" and can also can be used to manage operations if input in real-time. So, for example, the data warehouse might contain a table with individual service deliveries, another table of related clients, another table of related service providers, and a code table of service delivery types.
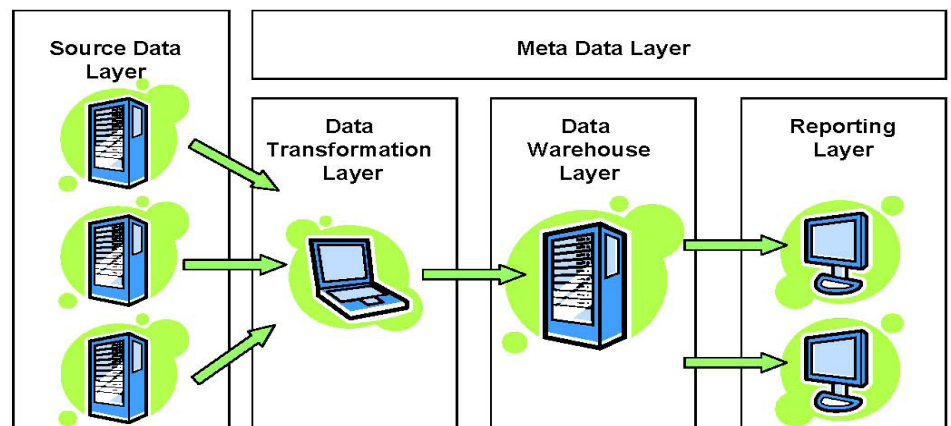


**Figure 2: Components of a data warehouse (adapted from Data Warehouse, 2006)**

Key to using data in a data warehouse is meta data. Meta data describe the meaning of data elements, including how they were cleaned and transformed in the data transformation stage.

# Data Warehousing Technology

# The Data Warehouse

The data warehouse tables are ready for use in many reporting tools and applications. On Line Analytical Processing (OLAP) reporting tools quickly provide information based on data cubes. One can choose the "slice," level of aggregation, criteria, and other settings for a query result. Data mining tools search for patterns in large volumes of data; for example, characteristics of youths who require services from multiple units. There is no limit to the applications of warehouse data. Of particular importance for many advanced applications is the operational data store, because of its flexibility.

The DHS Data Warehouse is a central repository of human services data to support decision-making. The operational data store is organized in a relational framework so that information about DHS programs and clients can be retrieved and analyzed. The DHS Data Warehouse contains more than 15 million client records, currently supplied from more than 10 independent operating systems, both internal and external to DHS. The Data Warehouse does not support real-time operation; rather the data is refreshed each month. Given the sensitivity of the data, access from the Internet is restricted. DHS staff members use a suite of analytical tools to connect to, extract, and analyze the data.

**History**
Prior to 1996, Allegheny County delivered a wide range of human services to its citizens through a network of independent county departments. Critics of the old system observed that while many individuals received services from several of those departments, there was little or no coordination of these services and no tracking of who was receiving services from multiple departments. In 1995, a blue ribbon panel, ComPAC 21, recommended that the existing human service departments be integrated into a single department. In 1996, the Allegheny County Commissioners responded by creating DHS.

Faced with the enormous task of reorganization, human services management staff from all disciplines were charged with drafting organizational plans for the new DHS. The newly appointed Director of DHS received these recommendations and invited a broad cross-section of private sector and provider agencies to add their expertise to the process. The goal was to achieve efficiencies by consolidating duplicated functions while preserving and enhancing program services and maintaining compliance with state requirements within a fiscally responsible environment.

# The Data Warehouse

It was soon recognized that integration posed very significant challenges for the new department's information technology systems. The formerly independent departments stored data on clients, providers, and services in more than 80 disparate databases and systems. The Information Systems Task Force was formed to review these complexities and formulate a plan of action.

Human Services Integration Fund

To assist with the development of the newly created DHS and fund DHS projects that fell outside rigid government funding streams, a coalition of 12 local foundations created the Human Services Integration Fund (HSIF), as an off-budget capacity-building fund. The Data Warehouse and many other management improvements have been funded largely by the HSIF. In 1999, with the backing of HSIF, DHS announced an RFP to develop a computing architecture to support the business process of an integrated DHS that would include eCAPS, the common client identifier operating application and the Data Warehouse, to integrate data from the separate program offices. Deloitte Consulting was selected as the contractor to work with DHS on the Information Integration Project.

In addition to supporting the original design of the Data Warehouse, HSIF member foundations questioned whether the Data Warehouse could also be used as a community resource tool for research, strategic planning, needs assessment, and program evaluation. Within a few months, an advisory committee was convened to research the answer. The committee consisted of an esteemed group of representatives from academia, the foundation community, human services providers, government entities, and the private sector. Through four subcommittees, this advisory committee identified the likely users of the Data Warehouse to be university-based researchers; health and human service agencies; policy and planning organizations including foundations, civic agencies and grant-making federations; and consumers of human services, their family members, and advocates.

Results of their research proved critical to the ultimate design of the DHS Data Warehouse. Their research determined categories of business questions important to the different groups of community stakeholders. Identified areas of interest included tracking consumers and aggregate demand for services; monitoring quality, cost, and aggregate outcomes; agency management and planning; tracking linkages within and between service delivery systems; and consumer choice and advocacy. The advisory committee also explored data integrity and confidentiality; marketing, education and technical support for users; and oversight and governance.

# The Data Warehouse

<u>Community Resource</u>

The DHS Data Warehouse has become a significant community resource. Most recently, in partnership with Three Rivers Connect, MAYA Designs, a group of approximately 20 community organizations, and the financial support of the Foundation Community, DHS has designed, developed, and implemented a web-based application for the residents of Allegheny County. HumanServices.net is a public facing Web site that contains up-to-date information on DHS-provided services as well as other community resources. The first phase of this project was placed into production in 2006. The next phase of the project will place considerable emphasis on the inclusion of community profiles. These profiles will describe both the demographics of and the provision of human services within each of these communities. This next phase will also continue the linking to other resource guides, the inclusion of health data (location and services of clinics, hospitals, WIC sites, etc.) and the development of an online interactive expert question/answer function.

**Data Holdings**

The data in the DHS Data Warehouse identifies and describes DHS clients. It contains identifying data such as social security number and gender, and descriptive data such as marital status and educational attainment. The DHS Data Warehouse matches this data with DHS-contracted providers and DHS programs serving these individuals, so that DHS can know who is receiving which services, and at what location.

The Data Warehouse further joins this data with data external to DHS, creating a multi-dimensional picture of its clients and the areas where they live, work and receive services. These external data sources include the Allegheny County Bureau of Corrections, the Pennsylvania Bureau of Probation and Parole, the Pennsylvania Department of Corrections, the Allegheny County and City of Pittsburgh Housing Authorities, and the Pennsylvania Department of Public Welfare. Data from these sources allow DHS to identify involvement and relationships between DHS and external populations, as well as compare the potential outside services available to them.

**Unique Features of Public Sector Data Warehouses**

Three unique features of public data warehouses are data agreements, lack of unique identifiers, and data privacy.

<u>Data Agreements</u>

While the public sector exists to provide public goods that are freely shared, bureaucratic red tape and "turf" make it difficult to obtain data from operational and external units. A major activity of the DHS Data Warehouse is identifying operational data stores and key personnel, assessing the contents of such data stores, and obtaining data supply agreements. DHS has some of the power of corporations, because it funds its units and service providers, but nevertheless increasingly finds that it must negotiate to obtain new data sources.

# The Data Warehouse

## Unique Identifiers

In the private sector, all entities that enter into transactions have unique identifiers, with accounts set up to ensure payment. In the public sector, personal identifiers may change for the same person. This is particularly pronounced for human services clients because of impairments and poverty. For example, much of the human services client population is transient, moving in the same or different neighborhoods. Without centralized registration, clients must often re-register for services as they move, changing spellings of names, addresses, dates of birth, and even social security numbers.

Any data that DHS obtains for its Data Warehouse must thus have several personal identifiers. In the data transformation stage, data analysts use rule-based algorithms to match several identifiers and partial identifiers. If enough match, a "hit" is declared and the person in question is matched up with his/her earlier records. Doing this "up front" greatly increases the utility of applications.

## Data Privacy

The other side of the unique identifier coin is data privacy. While aggregate data from data cubes is shareable, data from the more useful operational data stores describe individual persons and are private; for example, some of these data fall under HIPAA (Health Insurance Portability and Accountability Act of 1996) privacy requirements.

There are several steps needed to de-identify person-level data records. They are: (1) strip out names, (2) replace personal identifiers such as social security number with an arbitrary sequence number, (3) replace date of birth with age in years, and (4) replace street addresses with area identifiers such as census block or block group number.

# Applications

The immediate payoff from the DHS Data Warehouse was that data were integrated across its five operational units (see Figure 1). Before the availability of the data warehouse, little could be said about relationships between human services and clients across units. For example, managers assumed there was a significant overlap between clients accessing mental health and drug treatment services, or between mental retardation and mental illness—but had no confirmation or sense of magnitude. Likewise, managers had assumed that child protective services were the major entrée to services for children and their families. Managers also had little more than hunches about where geographically the majority of services were being delivered.

# Applications

Data from the Data Warehouse finally shed light on these issues. Analysts found that nearly half of drug and alcohol treatment clients also receive mental health treatment; conversely, only 12 percent of mental health clients also receive drug and alcohol treatment. Similarly, about one-third of mental retardation clients receive mental health treatment but only 4 percent of mental health clients receive mental retardation services. Contrary to the belief that child protective services are the primary entry point into the DHS system, the Data Warehouse showed that a child is almost twice as likely to enter DHS through the mental health system as through child protective services.

By geocoding client data, analysts were able to show where DHS clients live, allowing services to be located in appropriate places. For example, Figure 3 indicates the Pittsburgh neighborhoods where most child clients of DHS live (the Migration Analysis section of this document provides more information on facility location applications).

The DHS Data Warehouse is key to running a modern human services agency. It has also permitted more in-depth research and analysis. We selected a number of DHS Data Warehouse applications that illustrate the value and potential of this technology for advanced support of policy and management. Most of the applications represent ongoing efforts at various stages of completion or implementation.
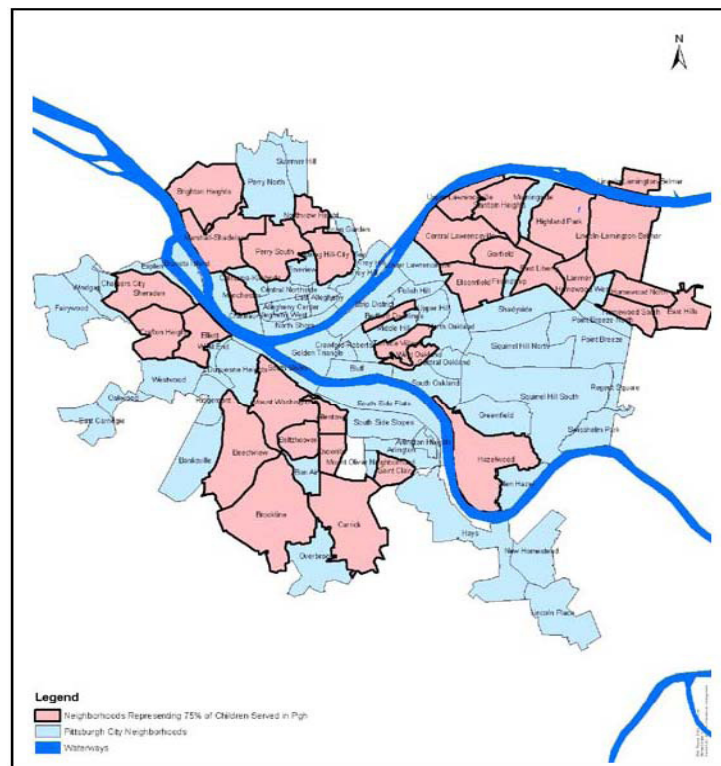


**Figure 3: Pittsburgh neighborhoods where 75% of DHS child clients live.**

# Applications

**Returning Prisoners' Behavior**

The full potential of the DHS Data Warehouse can be seen when data sets beyond those maintained by the Department of Human Services are added to the system. The application of this section is part of the work funded by Pittsburgh-area grant-making foundations to study the problems associated with the return of released prisoners back to society in Allegheny County.

Issue

Nearly all offenders sent to prison return to their communities — most within three years of imprisonment. If returning offenders are going to be successful after their release from prison, they must obtain essential human services such as drug and alcohol treatment, temporary housing, mental health services, and job training. To establish a baseline for research, the initial question asked was, "Do offenders returning from state prison use human services?"

Approach

To better understand the needs of returning offenders, DHS obtained data on Allegheny County commitments to the Pennsylvania State Department of Corrections. Data on 3,292 individuals committed from Allegheny County and released from State Department of Corrections (DOC) facilities from 2002 to 2004 were analyzed.

Data Warehouse Role

These data were integrated into the DHS Data Warehouse to determine the extent of service use by incarcerated persons. Sixty-six percent of individuals released from state prison to Allegheny County between 2002 and 2004 received human services as determined by matched records in the Data Warehouse.

Results

Analysis of the integration of DOC prisoners and the Department of Human Services yielded significant information:

• Of the 1,941 prisoners who accessed services, 41% received drug and alcohol treatment and 63% received mental health services.
• Of those DOC prisoners who received services from DHS, 68% received services before incarceration and 32% received services after incarceration.
• Of those receiving DHS services after incarceration, 19% received services within one week, 35% within one month, and 63% within three months of their incarceration.

# Applications

• • •

- Black DOC offenders are statistically less likely to access services from DHS than white offenders[1]. Male offenders are statistically less likely to access services from DHS than female offenders[2]. The likelihood of accessing DHS services increases significantly for both male and black prisoners following incarceration.

- 42% of DOC offenders who accessed DHS mental health services had no DOC mental health diagnosis. Of DOC the offenders with serious DOC mental health diagnoses, 87% received DHS mental health services.

- 52% of offenders accessing DHS drug and alcohol services had a Department of Corrections drug screen score of zero[3]. 77% of DOC offenders with a serious DOC drug and alcohol score access DHS drug and alcohol treatment services.

- DOC offenders who serve full sentences are statistically less likely to access DHS services than those who had been paroled. In contrast, re-parolees (those who have failed on parole at least once) are statistically more likely to access DHS services.

- DOC offenders with a high school education or more are statistically more likely to access DHS services.

**Policy Simulation Model**

In the 1970s, Alfred Blumstein, a criminologist at Carnegie Mellon University's Heinz School, built a policy flow model of the criminal justice system in Allegheny County, Pennsylvania called JUSSIM (Blumstein, 1980). It was an early example of modeling the interactions between multiple agencies to estimate system-wide impacts of policy changes on capacities and costs and benefits. A problem with such a model at the time was calibration with data: there were no databases, let alone data warehouses. Thus JUSSIM was a one-shot modeling attempt, with many approximations, and never became an ongoing policy tool. The DHS Data Warehouse makes it possible, today, to consider building and maintaining policy simulation models. Through a project-based course, graduate students in the Heinz School conducted a pilot study to build a JUSSIM-like model for driving-under-the-influence (DUI) crimes. This work was done as part of a grant awarded to DHS from the Allegheny County Funders in Criminal Justice, a working committee of six local grant-making foundations.

Issue

Often the responsibility of county governments involves a life cycle that moves cases through many phases and departments. An innovation in one department that increases the flow of clients into certain programs interacts with and has impacts on other programs and departments. Systemic questions arise as a result, include the following: Do impacted programs "downstream" of an innovation have the capacity to handle additional cases? Will policies made "upstream" change the flow of cases? Does the innovation improve costs and benefits relative

---

[1] Double sided p-value: 0.0238
[2] Double sided p-value: 0.0000
[3] A measure of addiction severity using the Texas Christian University Drug Screen II (TCUDSII); prior to 2001, this was assessed using the PACSI (an in-house tool). TCUDSII ranges from 0 (no addiction problem) to 9 (drug dependence). PACSI ranged from 0 to 10, with a similar interpretation. The database does not distinguish which tool was used, though.

# Applications

Pennsylvania has some of the most stringent DUI laws in the country, with criteria and penalties toughened as recently as 2004. Nevertheless, in 2004, 610 (41%) of Pennsylvania's 1,490 highway fatalities were alcohol-related and 546 (36%) fatalities involved drivers with blood alcohol content over the legal limit (Pennsylvania Drunk Driving Statistics, 2006). In 2005, the number of fatalities involving drivers with blood alcohol content over the legal limit increased to 559 *(U.S. DOT Releases 2005 Data on Alcohol-Related Traffic Fatalities, 2006)*.

In response, leaders in the Allegheny County Court of Common Pleas and its Probation unit designed and implemented a DUI Court in October 2005 for the worst repeat offenders. This is a community-based treatment program with court oversight. While there are many facets to this program, its main feature is that it enforces treatment for 16 months in most cases, with penalties including jail time for non-participation. There is some evidence that DUI courts reduce recidivism *(DWI/DUI Courts Work, 2006)*.

Many policy questions have arisen about the Allegheny County DUI Court. The initial implementation had strict and conservative eligibility rules. What if the rules were relaxed? How many cases would be eligible as a result? What is the impact on treatment facilities' capacity? What is the impact on jail capacity? What is the impact on DUI recidivism? Did the changes to Pennsylvania's DUI law in 2004 result in more or fewer cases?

Approach

Discrete event simulation of the simple flow model kind provides a system-wide model for answering questions. The approach is to build a flow chart of organizations and programs through which DUI cases flow, estimate the population of cases in each program per time period (year in this case), and identify the proportion of cases per period that depart each program to each downstream program. In flow modeling terms, each program is a node and the proportions of departing programs are branching ratios or probabilities. With a flowchart built and calibrated, it is possible to run scenarios through the system with changes in decision rules, program structure and capacity, etc. in successive years to estimate the long-term impacts of changes.

Figure 4 is a high-level flowchart of the DUI criminal justice system that represents the populations of the community and those apprehended for DUI violations, in pre-trial status, in court and disposition, and in supervision with a return to the community. Figure 5 is an example of an exploded flow model for supervision in Figure 4. The DUI Court sends cases to intermediate punishment, which has four levels of combined treatment and punishment.

# Applications

The total exploded flow model for DUIs has 24 nodes and 71 branching ratios to estimate. To evaluate DUI court, it will be necessary to drill down further and add more detailed/exploded program components. Thus, to maintain the flowchart model, it is necessary to build and maintain a model at the levels of Figures 4 and 5, and also to have the capability to expand nodes in more detail for special studies. With data available at the more exploded level, this approach is readily implemented using spreadsheet models.
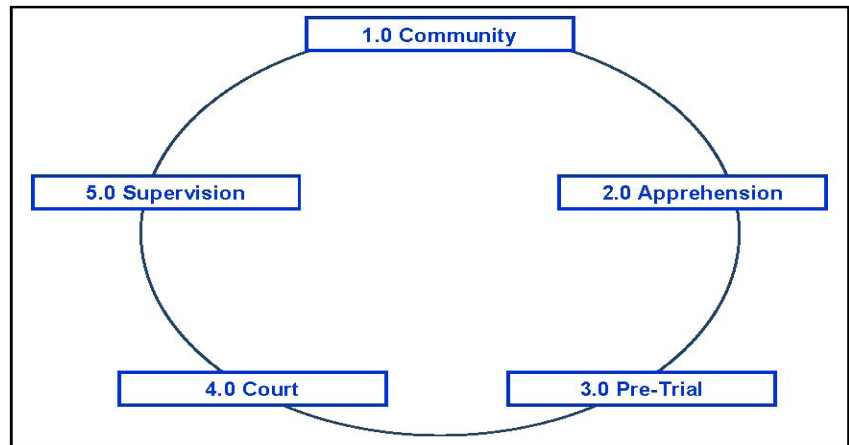


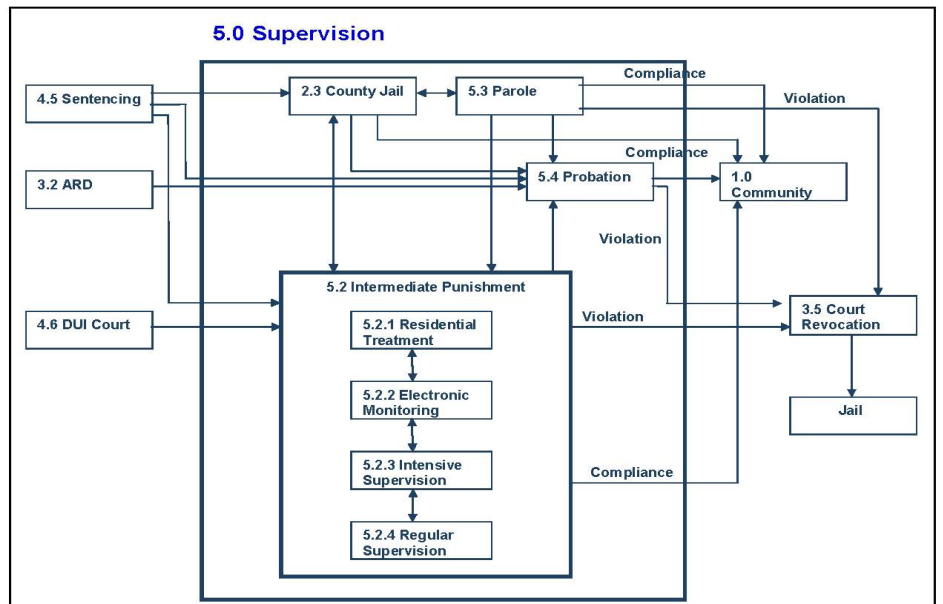**Figure 4: High-level DUI system flowchart**



**Figure 5: Exploded flowchart for DUI supervision**

# Applications

## Data Warehouse Role

The data needed to calibrate the DUI flow model were not yet in the DHS Data Warehouse, but the researchers were able to obtain 25,155 DUI records from the Allegheny County Court of Common Pleas and 27,969 from Adult Probation for 2004. These data were used in a pilot study and as a precursor to inclusion in the Data Warehouse. Data preparation and cleaning were documented for possible use in the Data Warehouse and included removing duplicate records, linking court records with probation records, and estimating node populations and branching ratios. The latter—estimating populations and branching ratios—required complex processing, but was captured as SQL queries that can be reused. Not all nodes of the full flowchart had available data, thus causing the flowchart to collapse to a smaller number of nodes (19) and branching ratios (66).

## Results

The student project was successful in demonstrating the feasibility of using flow models in policy making. Nevertheless, DHS does not have any experience at this point in actually applying such a model. At this stage, a policy flow model remains an attractive application of a county government data warehouse. It seems necessary for comprehensive policy analysis but requires significant new human resources for implementation. To build, maintain, and apply such models will require additional staff.

## Migration Analysis

Over time, populations tend to shift locations within a county, often due to economic development and demographic trends. For example, whites fled cities in the 1960s and relocated in suburbs, and over time there has been a migration of elderly populations from cities to suburbs.

## Issue

Policymakers need to understand how the location of their population is changing. School closings and openings should be based on the residence patterns of school-aged children and their families. Likewise, as the population moves and changes, so should the treatment centers, the polling places, and the supermarkets.

Policymakers may also want to know how policy changes, such as the demolition of public housing communities in urban cores, impact the surrounding communities. The data warehouse presents one of the few opportunities to study such patterns.

# Applications

<u>Approach</u>

Researchers at Carnegie Mellon University are using the DHS Data Warehouse to track the locations of former housing authority residents. The researchers have data on all former residents of housing authority communities that have been torn down; however, they don't know where the residents have moved to. Knowing that there is a significant overlap between individuals who reside in housing authority communities and those who utilize DHS services or Department of Public Welfare benefits, it makes sense to compare those individuals with records in the Data Warehouse to see what percentage match. The Data Warehouse has address histories for those individuals who match.

Another approach to modeling migration is to monitor a geographic area over time. For example, one could examine a public housing community before, during, and after demolition. In this example, we expect the total number of persons represented in the data to decline over time since some clients will move outside of the county, some will stop receiving services, and others will not be able to be matched with prior records. If studying juveniles, some will become adults.

<u>Data Warehouse Role</u>

The data warehouse is key to implementing both approaches of studying migration. In the first approach, data from the City of Pittsburgh Housing Authority are integrated in the DHS Data Warehouse and for those who match in the system, there will be current and previous address histories.

In the second approach, DHS client data (such as child welfare data) are examined in one or more specific locations for a determined time period before, during, and after an intervention (e.g., the demolition of the housing community) to determine the number and change of clients over time.

**Cost/Benefit Studies**

Cost studies, including cost effectiveness and cost/benefit analyses, are of critical importance to policymakers who must defend their investments.

<u>Issue</u>

In 2001, Allegheny County began a Mental Health Court to place mentally ill individuals in mental health services/treatment in lieu of incarceration. Several years after its successful implementation, policymakers were interested in understanding whether it was cost effective.

# Applications

## Approach

The RAND Corporation was selected to conduct the assessment. Initially, researchers planned to obtain informed consent from Mental Health Court participants in order to access their mental health and criminal justice records from state and county agencies. For each subject, researchers would construct a utilization record of mental health and criminal justice resources from two years prior to entering the Mental Health Court to the present. There is no comparison group – Mental Health Court clients would be compared to a counterfactual based on two assumptions. First, researchers would use sentencing guidelines from the district attorney's office to create the counterfactual criminal justice utilization that would have occurred if the Mental Health Court-brokered plea bargain had not taken place. Second, researchers would extrapolate the mental health utilization from before mental health court enrollment to give us a counterfactual mental health record. The fiscal impact of the court would be calculated as the difference between the cost of the actual utilization following Mental Health Court enrollment and the cost of the counterfactual utilization.

This approach had to be modified because many of the consents could not be obtained. For these individuals, criminal justice records were requested by the Department of Human Services, integrated into the Data Warehouse, combined with service datasets, and de-identified by DHS staff before being returned to RAND researchers.

## Data Warehouse Role

Without the Data Warehouse and the DHS staff members who manage it, the study may have been unsuccessful. Criminal justice data were integrated into the Data Warehouse and combined with client service histories. This analysis could not be conducted by RAND researchers because it would violate human subjects protections.

# Future Work

Future work of the DHS Data Warehouse centers on obtaining additional external data from related county agencies, building a real-time operational data store, establishing the family as the context for treatment, and building better meta data. With increasing stores of data from other county agencies, the DHS Data Warehouse has the potential to become the Allegheny County Data Warehouse, providing support for decision-making across many municipal agencies. Key to the real-time data effort is the development of a Master Client Index (MCI), which will allow for instant linkages between operational units. Additional linkages are needed to establish and track clients and their families. Finally, better meta data and access to meta data are needed to expand the direct use and understanding of data within the Warehouse.

# Future Work

**Development of a Master Client Index**

To provide DHS with a centralized intake process, DHS is designing, developing and implementing an MCI. This enterprise application will support the provision of integrated services to all DHS clients.

Creating a comprehensive view of clients across DHS is a difficult task due to the nature of many of the Department's existing information systems. These systems are unconnected data repositories that require special efforts to present the holistic view of clients needed to support policy decisions.

The MCI will help simplify policy decision-making by providing an integrated view of clients across DHS by:

• Providing one consistent and accurate identifier which will link all disparate identifiers,
• Standardizing client demographic data by providing one authoritative central reference point for key data,
• Providing advanced searching to ensure that all possible relevant clients are evaluated as potential matches,
• Helping maintain accurate demographic data by collecting updates from across DHS, and
• Laying the foundation for tighter integration of departmental systems, using a single identifier to allow for easier exchange of data.

**Service Delivery within a Family Context**

The addition of the Master Client Index function will allow DHS to integrate data on all the family members of an individual using DHS services. This will allow DHS to provide services to an individual within a family context and to view families as a unit. The relationship data will accommodate different definitions of families or households.

**Meta Data**

The DHS Data Warehouse specialists maintain detailed work logs for all efforts in the data transformation layer so that they can replicate and explain steps taken. The resultant operational data stores are the result of many decisions and transformation steps; in fact, only Data Warehouse specialists may create new information extracts. One resulting challenge, then, is to build up the meta data and access to it, in order to support manipulations and interpretations of the operational data stores and data cubes.

# References

**References**

Blumstein, A. (1980). Planning models for analytical evaluation, in Klein M & Teilman D (eds) *Handbook of criminal justice evaluation*. Thousand Oaks Ca: Sage.

*Data Warehouse* (2006). Internet Web page accessed 10/21/06: http://en.wikipedia.org/wiki/Data_warehouse.

*DATA WAREHOUSEI/DUI Courts Work* (2006). Internet Web page accessed 10/22/06: http://www2.potsdam.edu/hansondj/DrivingIssues/1083352479.html.

Gorr, W.L., M. Johnson, and S. Roehrig (2001), "Spatial Decision Support System for Home-Delivered Services," *Journal of Geographical Systems*, Vol. 3, pp. 181-197.

Inmon, W.H. (1995). "What is a Data Warehouse?", *Prism*, Volume 1, Number 1.

Johnson, M., W.L. Gorr, and S. Roehrig (2002), "Location/Allocation/Routing for Home-Delivered Meals Provision: Model and Solution Approaches," *International Journal of Industrial Engineering*, Special Issue on Facility Layout and Location, Vol. 9 No. 1, pp. 45 - 56.

Johnson, M., W.L. Gorr, and S. Roehrig (2005), "Location of Service Facilities for the Elderly," *Annals of Operations Research*, Vol. 136 (1).

Davis, A.K., A.M., Haley, M.T. Joshua, W.A. Maceira, J.L. Moon, C. Nesbitt, K.D. Schmidt, A. N. Sinha, and J.Yeh (2006), *Prisoner Reentry in Allegheny County: Geographic Information System and Policy Simulation Model*, Systems Synthesis Report, http://www.heinz.cmu.edu/systems/80.html.

Kimball, R. and M. Ross (2002). *The Data Warehouse Toolkit, Second Edition: The Complete Guide to Dimensional Modeling*. Wiley.

O'Hare, W. and M. Mather (2003). "The Growing Number of Kids in Severely Distressed Neighborhoods: Evidence from the 2000 Census", A Kids Count/Population Reference Bureau Report on Census 2000.

*Pennsylvania Drunk Driving Statistics* (2006). Internet Web page accessed 10/22/06: http://www.alcoholalert.com/drunk-driving-statistics-pennsylvania.html.

*U.S. DOT Releases 2005 Data on Alcohol-Related Traffic Fatalities* (2006), Internet Web Page accessed 10/22/06: http://communitydispatch.com/artman/publish/article_6064.shtml.