



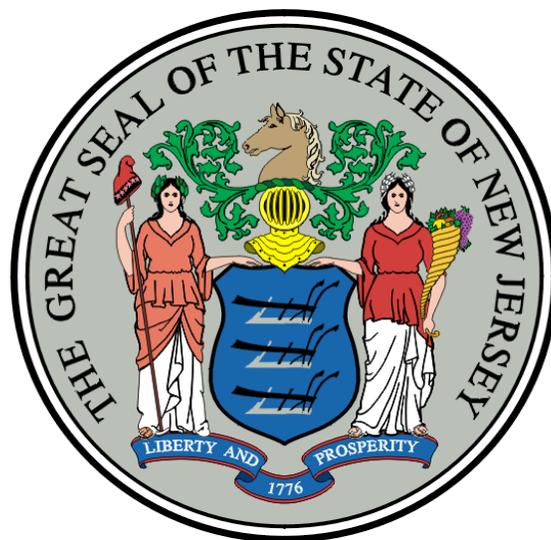
**2011 NASCIO Recognition Award Submission**

**New Jersey Office of Information Technology  
Office of Data Management Services**

**The New Jersey Enterprise Data Warehouse**

**Category: Data, Information, and Knowledge Management**

**State of New Jersey**



## **B. Executive Summary**

Throughout New Jersey over the last twenty years, agencies have struggled with providing integrated data reporting capabilities to meet management needs. Various approaches have been used, and all shared a few common weaknesses. Data definitions were not standardized, resulting in inconsistent data quality. Redundant efforts were spent on extracting and publishing data in multiple places and for multiple purposes without leveraging data, effort, or understanding. The business rules for the transformation of the data into useful information were locked in the heads of multiple, at times competing, analysts. The reporting infrastructure was expensive and did not provide all of the desired functionality. The data was often only available in the format required by one report community, and not in the format required by others.

In March, 2008, the Office of Data Management Services published version 4 of the New Jersey Common Information Architecture (NJCIA). This described a framework for developing an enterprise data warehousing (EDW) capability for the state to be used by multiple agencies across multiple subject areas. The framework prescribed a hub-and-spokes enterprise data warehousing environment to support all state agencies.

The objectives of the NJCIA were to create an enterprise data warehousing environment using robust tools that would create a store of standardized data that could be leveraged by multiple initiatives for operational reporting, analytical reporting, and interfaces to other systems. The environment would be supported by a data warehousing competency center that would be able to meet the diverse data warehousing needs of multiple state agencies.

Through December 2010, the original EDW has expanded to include many additional administrative data sources that are now made available to seventeen different agencies. From the original two subject areas, there are now fifteen. In addition to this administrative data, which we call the New Jersey Administrative Warehouse System (NJAWS), the EDW is also home to subject areas such as Criminal Justice Recidivism, Economic Growth, Pensions, Crash Records, EMS Responses, MV Revenue, Risk Management, Transportation Operations, Behavioral Health, and Property Management. It will soon be the home to multiple subject areas for Motor Vehicle Inspections, Children & Families, and Medicaid. But even this does not tell the entire story, as the EDW environment has been a significant part of many additional initiatives, providing data more quickly, more effectively, less expensively, and/or with more reusability. Because the EDW was already in place, Governor Chris Christie was able to roll out his comprehensive, data-driven government transparency site, **YOURMONEY.NJ.GOV** in only ten weeks.

### **C. Description of the Business Problem and Solution**

At its heart, the problems that New Jersey faced were data quality and data governance. Every time a manager, a policy maker, or a politician needed information, it was addressed by a one-off scramble to locate data, bring it together, and put it into a useful format. When looked at in isolation, the impact of any one of these efforts was not noticeable. Collectively, however, they resulted in hundreds of data files being moved throughout the organization in an unmanaged way, often with the same data be created and moved multiple times.

As a result, the state incurred substantial costs to support the data file creating. Worse, because there were no standards in place to govern the definition, transformation, or use of the data, the resulting information generated had inconsistent data quality. The legacy reporting tools available did not provide functionality that report consumers were demanding, and the tools required substantial technology support and custom programming each time a new report was required.

A number of obstacles stood in the way of addressing these data quality and support problems in a comprehensive way. After many years of keeping data internal to their organizations, a culture of data sharing was emerging, in which agencies reluctantly agreed to share data files between each other. While a step towards better information, this actually exacerbated the data quality problems as data was shipped haphazardly through organizations. Agencies resisted the idea of integrated data for reusability.

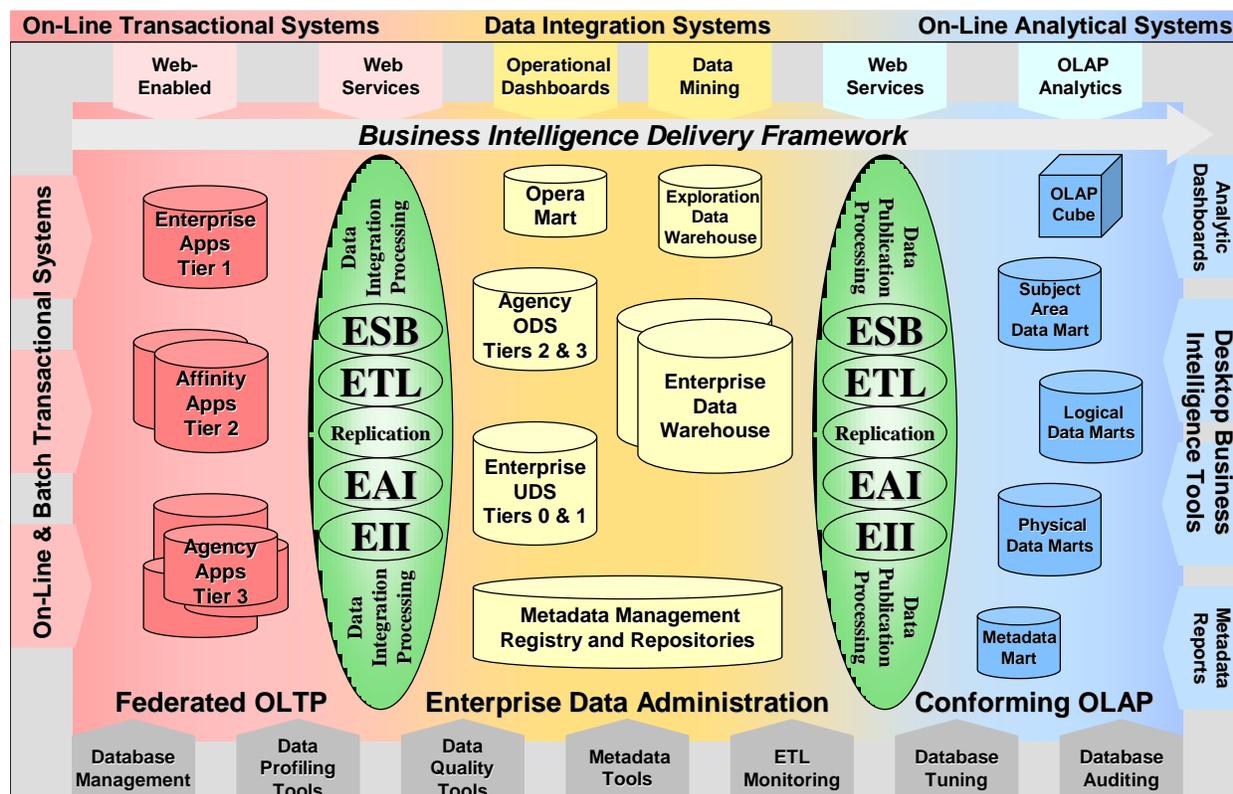
A second obstacle was that the reporting technologies at the time were often in the hands of analysts that were reluctant to give up their very prominent role in delivering information and solutions to management. A third obstacle was that the technology units supporting transactional systems were not comfortable with the idea of “their” data being available to others.

These obstacles were compounded by the lack of authority for a centralized approach. In 2007, the governor signed legislation designating the New Jersey Office of Technology as responsible for information technology in the executive branch. Shortly thereafter, the NJ Office of Legislative Services released an audit report that indicated that New Jersey needed to develop a strategic plan (architecture) for enterprise data warehousing.

As a result of these two events, the NJ Office of Data Management Services (ODMS) in the NJ Office of Information Technology (OIT) published version 4 of the New Jersey Common Information Architecture (NJCIA) in March, 2008. The NJCIA describes a framework for developing an enterprise data warehousing capability for the state to be used by multiple agencies across multiple subject areas. The framework prescribed a hub-and-spokes enterprise data warehousing environment to support all state agencies. Version 4 enhanced previous versions of the framework and also incorporated the governance directed by the new law as well as suggestions included in the audit report.

The NJCIA is based upon industry best practices as defined by NASCIO, by the Data Management Association (DAMA), and by the Data Warehousing Institute (TDWI).

## NJ Common Information Architecture Conceptual Model



Version 4.01

Revised March 2008

The objectives of the NJCIA are to create an enterprise data warehousing environment using robust tools that would be leveraged by multiple efforts for operational reporting, analytical reporting, and interfaces to other systems. The environment would be supported by a data warehousing competency center that would be able to meet the diverse data warehousing needs of multiple state agencies.

The ODMS was given the task of creating the enterprise data warehousing environment, standing up a data warehousing competency center to support all executive branch agencies, to migrate the existing reporting solutions into delivery systems based on the enterprise data warehouse, to develop policies and standards, and to evangelize this approach to agency business and technical staffs.

The first three projects brought into this environment were the New Jersey Administrative Warehouse System (NJAWS), the Crash Data Warehouse (CRASH-DW), and the Department of Transportation data warehouse (TransINFO).

The NJAWS re-engineering initiative began in early 2008 and was completed in October 2009. It integrated multiple independent data marts for financial, payroll, human resources, and other back-office systems into an enterprise environment that increased access to data, reduced storage requirements, improved performance, and streamlined administration. Initially an operational reporting area, this environment will eventually be a normalized, multi-subject area data resource that will insulate the state's data as legacy systems are retired and replaced with new solutions.

The CRASH-DW brings together data from State Police, Department of Transportation, Motor Vehicle Commission, and Department of Health and Senior Services for longitudinal crash analysis. This had never been done before. Furthermore, the initial release of CRASH-DW to the user community in September 2009 allowed, for the first time, full reporting capabilities across all the relevant crash data supplied by the four above-mentioned NJ agencies. Subsequent functionality was added in September 2010. Additionally, because of the enterprise approach, *non*-crash emergency medical service data is now available to the Department of Health and Senior Services for the analysis, and DHSS is able for the first time to meet its federal requirements to provide regular data to the National Emergency Management Information System, with no additional costs.

The NJ Department of Transportation had attempted to integrate its operations systems twice before over a period of five years, without moving beyond the planning stage either time. Obstacles included the lack of data warehousing technologies, the lack of a data warehousing methodology, and the inability to reach consensus on data definitions. The TransINFO project began in July 2008 and was implemented in June 2010. It overcame obstacles by taking advantage of the NJCIA data warehousing methodology and technologies, as well as using ODMS staff to create a sustainable data model and resolve the data definition conflicts. It leveraged roadway data that was initially part of the CRASH-DW, which resulted in significant savings.

Through December 2010, the original EDW has expanded to include many additional administrative data sources that are now made available to seventeen different agencies. From the original two subject areas, there are now fifteen. In addition to this administrative data, which we call the New Jersey Administrative Warehouse System (NJAWS), the EDW is also home to subject areas such as Criminal Justice Recidivism, Economic Growth, Pensions, Crash Records, EMS Responses, MV Revenue, Risk Management, Transportation Operations, Behavioral Health, and Property Management. It will soon be the home to multiple subject areas for Motor Vehicle Inspections, Children & Families, and Medicaid. It played a critical role in the success of Governor Christie's comprehensive government transparency site, **YOURMONEY.NJ.GOV**.

#### **D. Significance to the Improvement of the Operation of Government**

The enterprise data warehouse represents a fundamental transformation in the way that New Jersey state government views its data resource. Data is no longer viewed as the

property of a business unit, or worse, a business analyst. Instead, it is treated as a reusable resource. New Jersey recognizes the value of the six overarching drivers of the NJCDA – that data must be more accurate, more complete, more accessible, more reusable, as well as timelier and less expensive to administer.

## **E. Benefits of the Project**

This project addresses the NASCIO State Priorities for 2011, including in Category A the Nos. 1, 3, 5, 6, and 10.

While not originally the objectives of the NJ enterprise data warehousing program, the following initiatives have leveraged the environment to improve data quality, reduce development costs, and/or reduce development time.

- **Governor Christie's Transparency Site** – New Jersey's first comprehensive, government transparency center. YourMoney.NJ.Gov is designed to provide a clear picture of the state's revenues and spending for all citizens. The previous Open Public Records Act process was burdensome and expensive compared to Web-based search engines and data storage. This new site consolidates values of public information on one site that is easy to navigate. Its first release went live with two subject areas after only ten weeks by leveraging the EDW. Subsequently, four additional subject areas were added, and the scope was extended to include data from independent state authorities.
- **Payroll Integration** - A bi-directional data transfer process was quickly developed between the new statewide electronic Cost Accounting and Timekeeping System (eCATS) and the payroll mainframe payroll system to meet requirements, keep the project on schedule, and reduce development costs.
- **Employee Profile** - For the first time, the State has an integrated view of NJ employee information using data from six different legacy sources. This profile data is provided to numerous systems through the Enterprise Data Warehouse.
- **eOBCIS Global Justice XML Effort** – In partnership with the Department of Corrections, OIT has built a Global Justice XML-based data integration process that eliminates duplicate data entry of criminal justice information into two disparate systems, and leverages this integration for reporting and analysis. DOC has re-purposed several data entry staff while improving accuracy and timeliness. This is being expanded to include other criminal justice components. This effort is now being leveraged for an analytical reporting environment (data mart) to study recidivism rates.
- **Cancer Registry Data Quality Process** – Using the EDW data cleansing tools, data warehousing staff cleanses, standardizes, and geo-codes cancer registry information for the Department of Health and Senior Services to improve its data

accuracy. This service is now being provided to other DHSS units as well as other agencies, such as to cleanse heating assistance data for the Departments of Human Services and Community Affairs.

- **Universal Data Store** – What does County Code “02” represent? You would think this is a simple question with a simple answer, but it is not. The State uses a half-dozen different approaches to identifying counties in its systems. In some, “02” is Bergen County (the second county alphabetically). In others, “02” represents Atlantic County (the first county). In others, various alphanumeric codes are used (e.g. “A”, “AT”, “ATL”, etc.). This is just one example of the significant problem of integrating data from multiple sources. It is the reference data that enables integration, but it is the reference data that has the most problems with standardization and quality. The UDS is a section of the enterprise data warehouse that maintains this common reference data.
- **Data Feeds of dozens of data sources to dozens of systems** – On almost a weekly basis, the data warehouse teams at ODMS are contacted to obtain a data feed of data in the EDW. In the past, these requests would have gone back to the individual mainframe system teams so that a custom interface could be built.

There are notable benefits to not building custom interfaces. Individual requests for data add workload to stretched mainframe support teams and mainframe environments. They strain available processing windows. They add fragility to the overall system. They increase the replacement cost estimates when systems are considered for replacement. The development costs for mainframe extracts are significant.

In the EDW, the data has been populated by a single interface. The costs for deploying new data feeds out of this environment are substantially less – often by an order of magnitude – and the development time is quicker. In some cases, the data can be provided in a self-service format. If the mainframe system is ultimately replaced, the new system need only provide a single data feed into the EDW, and all of the data consumers continue to receive their data.

All of these efforts benefited from the use of components of New Jersey’s Enterprise Data Warehousing (EDW) environment, as guided by our New Jersey Common data Architecture (NJCDA). All of these efforts were completed leveraging existing resources. These efforts were completed more cost effectively, especially within the context of continuing support and maintenance costs. Many of these efforts were completed more quickly due to the existence of the EDW environment. In fact, some of these efforts could not have been completed at all without it.