# Application for Best Practices Recognition 2012

## Part 1 - Project Summary

**Project Title:** Louisiana HSRG Data Warehouse: One Version of the Truth

**Project Description (three sentences or less):** Since its inception in 1998, the main focus of the Highway Safety Research Group (HSRG) at Louisiana State University has been collecting and processing crash data. However, over the past few years, the HSRG has been focusing more of its effort and providing crash data that is more consistent, clean, and easily accessible. To support this new effort, the HSRG has designed and implemented a crash data warehouse and On-Line Analytical Processing (OLAP) cubes to produce a single version of the truth.

**Nominating Person Contact Information:**

| | |
|---|---|
| **Name:** | Mark Verret |
| **Title:** | System Administrator |
| **Agency and Office:** | LSU Highway Safety Research Group |
| **Address with City, State, ZIP:** | 3535 Nicholson Ext., Baton Rouge, LA 70803 |
| **Telephone:** | (225) 578-0283 |
| **FAX:** | (225) 578-0240 |
| **E-Mail:** | mark@lsu.edu |

**Lead Agency for Project:** Highway Safety Research Group (HSRG) - LSU

**Participating/Cooperating Agencies (if any):** Louisiana Department of Transportation and Development (LADOTD) and Louisiana Highway Safety Commission (LHSC)

**Which National Agenda goals apply?**

1. *Does it involve a leader(s) who promotes the importance of highway safety information systems, used for safety policy and program decision-making?*
   Representatives from the HSRG, LADOTD, LHSC and members from the Louisiana Traffic Records Coordinating Committee have been actively involved in the deciding what questions the data warehouse and OLAP cubes should address.
2. *Does it involve the coordination of the collection, management, and use of highway safety information among various organizations responsible for highway transportation policy?*
   The HSRG's data warehouse is the central repository for automobile crashes within the state Louisiana. The warehouse also contains other sources of information integrated with the crash data including; COBRA (breathalyzer) data, coroner report data, vital statistic data, driver license data, vehicle miles traveled data, and some roadway data.
3. *Does it represent an example of integrating the planning of highway safety programs with highway safety information systems*
   Information from the data warehouse and OLAP cubes have been used by the HSRG, LADOTD, LHSC, LA TRCC, NHTSA, FHWA, LA State Police, LA Sheriff Offices, LA

Police Departments, and other local & state agencies in Louisiana. The data warehouse and OLAP cubes have been designed to provide information for the LA Strategic Highway Safety Plan, NHTSA data reporting, GIS map development, hot spot analysis, and law enforcement reporting. Text mining projects are currently under development utilizing the data warehouse to analyze crash information to help detect fraud, helmet use, alcohol, and distracted driving.

4. *Does it represent an example where managers and users of highway safety information have utilized or were provided the necessary resources to select the appropriate technology to meet their information needs?*
The development of a centralized data warehouse allows for consistent reporting and analysis to be performed by internal and external users. Data dissemination is made available through static reports, web sites, interactive web tools, Excel applications, and other methods as needed.

5. *Does it represent examples of highway safety professionals being trained in the analytic methods appropriate for evaluation of highway safety information?*
In person training, training videos, and documentation is provided with each specific application to ensure it is utilized properly and to its full potential.

6. *Does it involve the promotion and use of technical standards for characteristics of highway safety information systems, critical to the development and management of highway transportation safety programs and policies?*
The state of Louisiana has an established standardized collision reporting form based primarily on elements used in reporting to federal transportation agencies. A majority of the Model Minimum Uniform Crash Criteria (MMUCC) reporting guidelines are utilized. The data collected adhering to these criteria is provided to help reporting agencies improve their data collection efforts.

**Which steps in the management process does the project support?**
- Establish Safety Goals –
Each year the Louisiana Highway Safety Research Group establishes goals to help support state traffic management efforts. Goals for 2012 include:
  o Decrease the total number of injury/fatal collisions.
  o Increase the number of agencies utilizing electronic reporting.
  o Improve overall data quality and reporting metrics.
- Identify Problems –
  o Decrease the total number of injury/fatal collisions:
    - Improve identification of problem areas in a timely manner to allow for effective corrective actions.
    - Improve identification of contributing factors to allow for appropriate corrective actions.
  o Increase the number of agencies utilizing electronic reporting:
    - Reduce the time in which collision reports are reported to state and federal agencies by having originating agencies utilize electronic methods.
    - Interface with third party applications to facilitate electronic data acquisition.
  o Improve overall data quality and reporting metrics:

- ▪ Providing feedback to local law enforcement agencies allows them to address potential issues at the point of origin.
- • Plan Programs/Counter Measures –
To meet the stated goals the HSRG meets periodically with participating agencies and hosts an annual meeting with all law enforcement agencies in the state to provide presentations and receive feedback. The data warehouse and OLAP cubes are designed to help identify problem areas, allowing agencies to then create the proper counter measures.
- • Implement Programs –
The data warehouse will serve as the basis of future efforts to improve the quality, consistency, and accessibility of crash information. Connecting data from various sources in a singular system that can then be utilized for reporting and analysis will help provide a broader understanding of the problems regarding highway safety.
- • Monitor Program Operations –
Capturing web statistics, obtaining user feedback, and focusing on particular improvements will be used to demonstrate how well the data warehouse and OLAB cubes are performing.
- • Evaluate Effectiveness –
Effectiveness will be measured on the usage of the system and the effect it has on meeting the goals presented.

**Reference the priority in your traffic records strategic plan to which this project applies:**
**Project Cost:** planned $: 250,000 actual $: 250,000

**Extent of Project Implementation:** The data warehouse and a few OLAP cubes are currently in production. The HSRG has created many tools (websites, static reports, ad-hoc reports, and excel spreadsheets) that are available and being utilized by many agencies throughout the state. The HSRG is always working to develop more tools to help integrate additional data sources into the data warehouse and create new tools to help make the data more accessible, visual, and meaningful to decision makers.

**Summary of Project Benefits: What was improved, who benefited, and how?** Benefits have been realized through the development of this project, even from the earliest stages. By having the data in a warehouse environment a richer set of analysis and reporting is able to be implemented. Data from disparate sources can be related and standardized. The HSRG and the agencies we serve all benefit from the enhanced ability to provide information faster and with less complexity.

OLAP cubes created from the data warehouse allow end users to easily obtain data without having to understand the back end data structure, join tables, and know how to write SQL statements. Users can easily drag and drop facts and dimensions allowing them to 'slice and dice' the data to obtain meaningful information.

## Part Two:  Project Detail

**Project Description:**

Companies have collected transactional level data for over the past five decades.  This data, alone with other sources of internal and external data, have resulted in a large collection of information in various sources for many companies. The ability of managers, at any level, to access, analyze, and interpret relevant information from these large data sources, are becoming more crucial to the success of organizations.  Many companies are facing tough economic times and are relying on fewer employees not only to operate, but grow the organization.  With the advancement of technology and software applications, mangers have the ability to utilize their desktop or laptop computers to obtain relevant, meaningful, and timely information.  Managers can now utilize their company's data sources to make effective decisions based on relevant information, not just personal instincts.

Organizations first used computers in the 1960s for transaction processing.   Businesses perform multiple transactions on a daily basis, for instance, creating a sales order, making a payment, receiving inventory, hiring an employee, or performing payroll. Data collected from these transaction processes are called transactional data and can be defined as information obtained by an organization to capture a business transaction.

Since the cost of computer storage has dramatically decreased over the past decades, many companies have elected to capture and store transactional data using databases called online transaction processing (OLTP) systems.  These OLTP systems are used to record the business transactions as they happen and the collection of all the transactional data serves as a history of the company's performance over time.

While these OLTP systems contain a vast amount of information, they are not ideally structured for aggregation of figures, development of reports, and ad-hoc querying by business managers.  OLTP systems were designed for collecting and storing transactional data, not enriched analysis and reporting of information.  For these reasons, mangers cannot easily utilize the data stored in OLTP systems to assist them in making effective decisions.

To help overcome these issues, a specific set of BI applications can be used to extract, transform and load (ETL) the transactional data into a data warehouse.  Extraction is used to pull data from multiple data sources, such as OLTP systems, data files, websites, and application systems. Transformation techniques are utilized to clean, standardized, and manipulate the extracted data.  Cleaning the data occurs when data quality checks are performed to validate that the data is complete and accurate.  Standardizing the data means ensuring the data formats are uniform, code values are constant, and unique identifiers are consistent.  Manipulating the data entails: (1) merging data fields, such and first and last name, (2) computing values, such as number of fatal crashes, and (3) altering vales, such as assigning 'M' for males and 'F' for females.  Finally, the new transformed data is loading into a data warehouse (DW).

The purpose of a DW is to assist with analyzing data to report information.  The DW is a separate database and has a different data structure from the traditional OLTP system.  It contains data from multiple sources and contains data which has been cleaned, standardized, and transformed.  This makes the DW a much better tool for analysis and reporting of data over the traditional OLTP systems.

The DW, if designed appropriately, and utilizes the proper BI tools, can then be made into a multidimensional database for data analysis and reporting.  This new data structure, called online analytical processing (OLAP), enables users to interactively analyze multidimensional data from multiple perspectives.

The goals of the project are:
- Provide a simple mechanism for decision makers to answer complex problems
- Offer the ability for ad-hoc reporting and data-mining
- Create a single source for all data queries and all reports
- Use out-of-the-box solutions to reduce development time and costs

The HSRG data warehouse was completed on-time and on budget.  The functionality provided has also met expectations and continuous improvements are implemented as needed.

**Referring to the National Agenda Goals, tell how your project relates to each one you listed in Part One of this application:**

1. *Does it involve a leader(s) who promotes the importance of highway safety information systems, used for safety policy and program decision-making?* The data warehouse is collaborative product that has evolved from the data needs of state and local agencies.  Current and anticipated data needs have driven the development of the system.
2. *Does it involve the coordination of the collection, management, and use of highway safety information among various organizations responsible for highway transportation policy?*  The data warehouse is designed to effectively allow for the management and dissemination of statewide crash and crash related data.  The system allows for authorized users to access data relevant for their particular needs.  Providing targeted information reduces the time and effort expended to get to what they need.
3. *Does it represent an example of integrating the planning of highway safety programs with highway safety information systems?*  Once the data is collected, the portal allows for presenting the data in virtually any possible format to make the information useful for decision making.  Pre-defined reports are currently being generated on a nightly basis.  The portal will add to these reports by allowing for more flexibility and use of "digital dashboards" to enhance decision making.
4. *Does it represent an example where managers and users of highway safety information have utilized or were provided the necessary resources to select the appropriate technology to meet their information needs?*  The data warehouse allows for any authorized user to run analysis, access pre-defined aggregations, and correlate data in new ways via ad hoc mechanisms
5. *Does it represent examples of highway safety professionals being trained in the analytic methods appropriate for evaluation of highway safety information?*  Training documents

and videos are available to help agencies better utilize the applications and provide options they may not have been aware of before.

6. *Does it involve the promotion and use of technical standards for characteristics of highway safety information systems, critical to the development and management of highway transportation safety programs and policies?* The core set of functionality is based on the statistical reporting provided to the federal agencies.

**Describe the major process steps for your project, including any unique aspects that enhanced success:**
We decided to use commodity based products to help keep the project on time and on budget. This also helps in the event technical support becomes necessary during the development of the system. The specifications and utility of the system are taken from agency requests and government mandates. This ensures the system meets the needs for official reporting as well as individual agency decision making needs.

**Provide the evidence and reasoning used to determine the success of the project:**
Since the data warehouse became available not only has standard data analysis become accessible, but the time needed to fulfill data requests has greatly reduced. Prior to the warehouse getting data for some ad hoc requests required complex queries to join data from various sources over multiple years. Now tools have been developed to literally provide drag-and-drop capabilities to find answers to even complex questions.

**Why should this project be recognized as a best practice in traffic records?**
Data collection is well established and common-place whether it is from local to state, state to federal, or local to federal. This information more often than not flows one way. Providing information back to the originating agency is increasingly important to allow agencies to:
   a) Make more informed decisions when deciding where to allocate resources.
   b) Plan and anticipate more accurately for future needs.
   c) More effectively make data quality improvements.

The HSRG's data warehouse provides tools that allow agencies to get the information they want and need in a timely manner to facilitate making effecting business decisions. Along with current systems and information, the warehouse is designed to allow for the addition of future data from systems that may come on-line at a later date.

As far as we know, the HSRG is the first agency in the country to develop a data warehouse and OLAP cubes to focus on crash data analysis. We have shown our products to many local, state, and federal (both private and public) agencies, all of which have been very impressed with our efforts.

*Please take a look at a tool that is currently under development to help agencies more easily analyze the crash data.*

**Help Video:**
http://datareportsdev.lsu.edu/Resources/Videos/Overview/SWF/ReportBuilder.Overview.Video.html

**Interactive tool:**
http://datareportsdev.lsu.edu/ReportBuilder.aspx