.

Radical BI

at

Lawrence Livermore National Laboratory

TDWI Best Practices Application

for

Radical Business Intelligence and Data Warehousing

March 30, 2007

Prepared by: Dave Biggers, ERW Production Manager,

Enterprise Architecture

and Data Provisioning

Table of Contents

tdwi

2007 Best Practices in Data Warehousing Awards
Leading Innovations in Business Intelligence & Data Warehousing

Deadline: April 2, 2007

* Required fields.

*Name of Nominated Company	Lawrence Livermore National Laboratory
*To which category are you applying? *You must submit a separate application for each category to which you are applying. Limit is two categories per organization.*	**Check one below:** ☐ Enterprise Data Warehousing ☐ Customer Intelligence ☐ Dashboards and Scorecards ☐ Enterprise Business Intelligence (EBI) ☐ Operational BI ☐ Data Governance ☐ Master Data Management ☐ Unstructured Data Integrated with BI ☐ BI/DW On A Limited Budget ☐ Predictive Analytics ☒ Radical BI ☐ Government and Non-Profit
*Lead Business Sponsor or Driver at Nominee's Firm Name, Title, Phone, E-mail, and Role	**Tom Brengle, ICS Department Head 925-422-8505, brengle1@llnlg.gov, Enterprise Sponsor**
*Signature and Date	
*Lead I.T. Contact at Nominee's Firm (must be different person than above) Name, Title, Phone, E-mail, and Role	**Michael Piscotty, IT Manager, Enterprise Architecture and Data Provisioning, 925-423-9109, piscotty1@llnl.gov, Enterprise Reporting Workbench Project Manager**
*Signature and Date	
Contact at Solution Sponsor's Company (If Applicable) Name, Title, Company, Phone, E-mail	N/A
*Signature and Date	

Note: Your application is not considered complete until you print, sign, and **fax back this page to TDWI at 425-687-2842. The information contained in this application is used solely for the purpose of selecting winners for the Best Practices program but is otherwise considered confidential by TDWI staff and judges. If you are selected as the Best Practices winner, your signature authorizes TDWI to promote your organization in TDWI public relations and marketing efforts.*

A. BACKGROUND

Company Description. Describe your (the Nominee's) company in one paragraph.	Lawrence Livermore National Laboratory (LLNL) is a premier applied science laboratory that is part of the National Nuclear Security Administration within the Department of Energy. LLNL is responsible for ensuring that the nation's nuclear weapons remain safe, secure, and reliable through application of advances in science and technology.

B. SHORT QUESTIONS - Respond to all questions below.

Other Contests. Has this project been submitted to other contests? If so, which ones and when?	No
Rollout Date. What month and year did the system being nominated officially go into production?	Initial implementation of a Data Warehouse solution was in 1986. The latest iteration known as the Enterprise Reporting Workbench (ERW) was rolled out in October of 2005
Active Users. How many business users use the system at least once a week?	1250

Types of Users. What percentage of the users fall into the following categories?	80% Casual Users (View reports several times a week)
	20% Power Users (Explore data regularly)
	   % Customers/Suppliers
	   % Other (Please specify): ____
	
	100%

Source Systems. What number of distinct source system applications does the data warehouse draw from?	Enter an Integer: **51** *AMS, FACTS, Effort, Ledger RMS, Human Resources, Travel, TID, Procurement, Electronic Order, Facility Cost, People, Configuration Management, Facility, Plant Services, SEP, Internal ICS, etc.*

Source System Percentages. What percentage of data in the warehouse comes from the following sources?	% Mainframe or minicomputer
	95% Relational
	 3% Desktop
	 2% Other files
	   % External data
	   % Other. Please specify: _____

Load/Update Intervals. What percentage of data is loaded in the following intervals?	% Quarterly 20% Monthly 50% Weekly – Fact Data 30% Daily – Dimensions and Operational Data % Less than daily. Please specify the interval and update mechanism: _____
Data Volume. How much data is in the data warehouse, including any downstream data marts or operational data stores? Express this in whatever terms you commonly use, like records, tables, files, gigabytes, terabytes, etc.	180 Gigabytes

What is the **2007 maintenance** budget of your system? (Please put a check the correct range at right):	☐ Less than $100,000 ☐ $100,000 to $500,000 ☒ $500,000 to $1 million ☐ $1 million to $2.5 million ☐ $2.5 million to $5 million ☐ $5 million to $10 million ☐ $10 million +
What is the **2007 capital** budget of your system? (Please check the correct range at right):	☒ Less than $100,000 ☐ $100,000 to $500,000 ☐ $500,000 to $1 million ☐ $1 million to $2.5 million ☐ $2.5 million to $5 million ☐ $5 million to $10 million ☐ $10 million+
Team. How many full-time equivalent staff are on the current BI team, including external consultants and contractors? What percentage is external to the company?	6.5 Number of FTEs on Current Team 0% of External Consultants/Contractors in above
Roles. How many FTE staff fill the following roles? (Include external consultants in your FTE count.)	0.05 Business sponsors/drivers 0.5 Project managers 1.5 BI architects/developers 2.0 ETL architects/developers 0.5 Data architects/modelers 0.5 Subject matter expert/business analyst 0.0 Business requirements analyst 0.0 Data modelers 0.5 DBAs 0.5 DW Administrators 0.5 Trainers Other notable roles with more than one FTE:
Initial Roll Out. Please indicate how long it took to roll out the initial system, the total cost to roll out that system, and the years until payback.	**Time** (from approval to initial roll out): 2 years **Start Date: 10/1/03 End Date: 10/01/05** **Cost** (including HW, SW, Services, Labor): $850,000 (Technology Migration) **Years** Until Payback (or estimate): immediate, DW users became ERW users immediately upon ERW implementation.
Executive Perception. Which **best** describes how your top executives view the BI system? (Select one.)	☐ Analytical tools that empower knowledge workers ☐ Monitoring system that reduces costs, boosts efficiency ☐ Mission critical system that drives processes & profits ☒ Strategic system that provides a competitive advantage
Stewardship/Governance. Describe the steering committee(s) or person(s) that set direction for the system. One paragraph total.	The Enterprise Reporting Workbench application efforts are sponsored by the CIO and directed by the ICS Department Head and Enterprise Services APL. The Enterprise Data Depot, data repository and provisioning service, is governed through partnerships with the ERW/EDD support team and the business unit data providers. Adding new business entities to the EDD is the financial responsibility of the business data supplier (owner of the business application).

C. Best Practices Essay
Radical BI at Lawrence Livermore National Laboratory (LLNL)

In 1985, before there were *prevailing* practices and *off-the-shelf* information retrieval solutions, LLNL developed a data warehouse (known as *ASSIST*) for an IBM mainframe running VM/CMS and a Nomad 4GL environment. Through three iterations and architectures, it is now, a portal-based Java and Oracle RDBMS application considered highly successful in delivering accurate and timely information to the Laboratory.

While data warehouses are now common place, several of our approaches and innovations stand out in today's data warehousing environment. We offer four such innovations that combine to give our users an uncommon level of capability and control in addressing their information needs. For context, each will be described individually before making a comprehensive case for recognition. As they are discussed, please keep in mind that unless noted, they were introduced during or before the '90s.

Modular Architecture –
Over the years, from *ASSIST* to *Data Warehouse* (DW, 1998) to *Enterprise Reporting Workbench* (ERW, 2005), a modular architecture has been employed and continually leveraged to allow users an extremely high level of flexibility, control and self-sufficiency in retrieving and/or disseminating business information. For instance, common report output comes from a *Report Definition* that is made up of independent components, a *Report Format* and a *Report Filter*. The ERW interface makes it easy for a user to create formats and filters (or use *Enterprise* versions of either) and link them to create *Reports*. Any format can be used with multiple filters and vice-versa, and all filter conditions can be set static or prompted at runtime. Finally, these *Reports* can be run immediately and/or saved for later use.

As Report Definitions are made up of components, they become components in the greater information delivery process. The ERW has a *Batch* module that automatically schedules, runs and delivers 15,000 reports each month based on the calendar and our business data load cycles. Users create *Batch Groups* consisting of one or more *Reports* that make sense for his/her needs (e.g. a cost report, a staffing report and a procurements status report). That *Group* (now, itself, a defined component) can be used in one or more *Batch Jobs* that will automatically run every time the relevant data sources are refreshed, usually weekly. Each *Job* contains its own filter criteria (via the prompted filters noted above) and specific distribution instructions, including the recipient(s) and his/her preferred output format - xls, pdf or html.

With the ERW, we expanded on our modularity to incorporate the ability to interact with commercial packages such as Business Objects or Oracle Reports. This evolution has lead to the *Workbench* reference in the name – a place where reporting tools can neatly coexist. Now lookups, graphical portlets, drill down facilities, ad hoc and scheduled reporting, excel integration, remote reporting services and remote data sources are all available from the ERW Home Page, a comprehensive workbench of enterprise reporting tools.

Virtual Dimensioning –
From inception through ERW, the overall data model had relied heavily on an enhanced star schema architecture. Aside from common institutional dimensions built around accounting and organizational structures, people, and facilities, (locally known as *Institutional Reporting Attributes* or *IRA*s), users can create and use their own virtual dimensions (*User Reporting Attributes* or *URA*s) around the same *base attributes*, account, person and facility numbers. All IRAs and URAs are available for filtering, sorting and display in any report from a business entity that contains that base attribute. IRAs are refreshed daily, and URAs are fully under the control of that user.

With ERW, application specific dimensions (*Application Reporting Attributes* or *ARAs*) were deployed, but some providers were opposed to wide display or use of their ARAs for business sensitivity reasons. Now, provider applications have the ability to control access to such attributes based on user role and identity.

Finally, the ERW's dimensioning architecture allows for rapid deployment of whole new dimension sets without programming or the need to create new view structures. Given that the dimension data is available, full deployment can be completed within minutes by simply adding or changing metadata.

(Virtual Data Dimensions is a major part of our TDWI Evening Session presentation in Boston, in May, 2007)

Radical BI at Lawrence Livermore National Laboratory (LLNL)

Desktop Integration –
Over the years, common desktop skills and capabilities have grown and become an integral part of today's retrieval and analytical processes. In the mid '90s, we found some users were re-keying data from paper reports into emails and other desktop processes so we preserved and staged their report files for FTP download. This service enabled users to more easily and accurately include report data in subsequent desktop activities. (Today, all output is electronic and available in full-featured xls, pdf, tsv and html.)

Early versions of user virtual dimensioning (above) allowed users to create and manage their URAs through an online interface. Driven by increased user skills and preferences towards spreadsheets (e.g. Excel), we deployed an ability to export and import URAs in spreadsheet form. This allowed users to edit their URAs online or download them to Excel for more major maintenance.

Last year, we deployed a new capability that (sort of) reverses the traditional flow of integration. Users can now, integrate post-retrieval analysis with report processing within the ERW prior to distribution. Over the years, we've seen numerous users create specialized reports designed for download into Excel. Often, they would combine several reports into a single workbook and then build pivot tables and charts, merge information and create summarization sheets, etc., before distributing the resulting workbook to co-workers and/or management. To reverse that flow, our new capability, the *Desktop Data Integrator* (DDI), allows users to create that workbook once and upload it to ERW where its data sheets and all pivots and summaries sheets will be automatically refreshed with each reporting cycle. The completed workbook can then, be distributed to its final recipient(s) or back to the authoring user or both via the ERW's report web and email based report distribution mechanisms.

Open and Enabling Philosophy –
LLNL's data warehouse began as a mainframe application designed to integrate and make available disparate business data from a multitude of corporate applications. The reporting tools and capabilities satisfied the needs for the vast majority of the business data focused population, but for a few, there were additional requirements or alternative preferences. Rather than view these as potential threats to our viability, our philosophy has been to encourage, recognize and support alternative approaches. When it became apparent users were looking to include report data in further processing activities, we added tsv and csv output formats to our user-selectable options, and we created a *Web Repository* to stage all reports for easy electronic retrieval if desired.

When small desktop systems began to proliferate (~2000), many required common institutional data such as account structures, people data, building directories, etc. We created and maintained the *Institutional Data Retrieval Facility* as an early web-enabled service so those users could easily retrieve current tsv, csv or xml versions of that data.

When some users and organizations began running global ERW reports for download into their local systems, we created the *Enterprise Data Depot* (EDD), a set of Oracle views cloned from the DW reporting data sources with access controls. These EDD views are directly accessible to the LLNL community via ODBC and other forms of connectivity. We also, added "SQL" as an output option from DW ad hoc reports to help users get started. If they want, users can run any ad hoc report to SQL and use that code as a model for direct query. This EDD facility has leveraged user creativity and become a source of institutional data for several processes that ultimately create data that ends up back in ERW and EDD.

There have been other approaches and innovations along the way, but these four seem to be at the core of 20 years of successful data warehousing. The balance of this discussion will focus on correlating our approach and innovations to the evaluative criteria.

Business Impact – *What is the business value of the DW/BI initiative?*
By 1987, LLNL *had* a comprehensive administrative data warehouse (ASSIST) that integrated business data from all traditional (and disparate) business operations, e.g. finance, procurement, human resources, business operations, etc. The foundation for today was set. The data was correlated and aligned and available from a single source via IBM 3270-style online queries or through paper reports from a user-managed, scheduled *Batch* process. Broad scale exposure to correlated data had a telling (and positive) effect on the quality of the source

applications. And through better monitoring capabilities, ASSIST directly contributed to better decision making, lower costs and overall improved financial management.

ASSIST, though modular from a user perspective, was a wholly encased application which proved troublesome in the mid '90s as the original technologies grew obsolete. We could not easily *evolve* to the newer platforms and software because of a total inter-dependence within the ASSIST environment. We had to replace everything to move forward to DW in 1997. This was costly, and it hammered home the value of modularity. By separating data stewardship from application infrastructure and user interface software in the DW, we had a much smoother and less costly transition to the ERW and ended up with a much better application, in part because we did not have to re-architect the EDD, the data stewardship component.

From a user's perspective, virtual dimensioning and especially URAs, allow our users the ability to model the Lab as they *see* it with a minimum of effort. Current day project management software with WBS capabilities is starting to offer that flexibility, but DW-ERW users have been describing their data worlds for many years. Coupled with our early recognition that the desktop is where the users work, they've also been able to use the tools of choice (Excel) to manage those configurations … and re-configurations.

From the '90s and the *Web Repository* to today with the DDI where the workbook is designed locally, uploaded and then automatically refreshed and distributed, we have always focused on making it as seamless (or at least painless) as possible for a user to integrate institutional information into their local work processes.

All of the foregoing in some way, relate to openness and enabling of users, but possibly the biggest impact derived from *openness* is the advent of the EDD facility. EDD access has spawned many *home grown* processes and mini-apps, but one in particular has revolutionized the way most LLNL employees manage their resources. One organization created an extremely innovative web-based cost and effort tracking system complete with charts and drills in a spreadsheet style interface. It became very popular throughout LLNL but impossible to extend and support in its original incarnation.   ERW then, partnered with the authoring organization to rewrite and centrally support it. Now, it is part of our suit of tools and the most widely used online retrieval system at the Lab … the backbone of our financial management tool set.

Maturity – *To what degree has the solution's "vision" been implemented?*
LLNL's DW maturity is already, well documented, but one point bears emphasis: While our original *visions* have been realized, our visions were not static. They evolved with emerging technologies and capabilities so we were (… and are) continually in quest.

Relevance – *Does the BI/DW environment exemplify best practices that other companies can adopt?*
The architectural attributes in this submission, Modular Architecture, Virtual Dimensioning and Desktop Integration coupled with our philosophy are great objectives and have huge payoffs if achieved, but a home grown approach may not be right for everyone. We started before there was a data warehouse, and this is how we've evolved. Notwithstanding, modularity, virtual dimensioning and integration are attributes worth pursuing, in all DW implementations, whether purchased or home grown.

With respect to our philosophy, our openness and focus on enabling users is our shield against becoming obsolete. Without that focus, the community will eventually move on.

Innovation - *Does the BI/DW use an innovative design or approach?*
LLNL's data warehousing efforts have a history of innovation and discovery. While many of the advanced capabilities have become mainstream, it was all new at the time it was released. DDI is our newest innovation and very much ahead of the capability curve, but other innovations will emerge as our visions move forward.

This presentation promotes several attributes of Lawrence Livermore National Laboratory's business data warehouse. In 20 years, we've done a lot of things right and some things wrong. In balance, we believe the component-based modular architecture, virtual dimensioning, DDI, and our "Reporting Workbench" concept are worthy of consideration as a best practice for the Radical Business Intelligence and Data Warehousing category.