

# Conference Trip Report

## 2004 Wilshire Meta-Data Conference & DAMA International Symposium

*Century Plaza Hotel, Los Angeles, California • May 2-6, 2004*

Compiled and edited by Tony Shaw, Program Chair, Wilshire Conferences, Inc.



This report is provided as a service of Wilshire Conferences, Inc.

The 2004 Wilshire Meta-Data Conference and DAMA International Symposium was held before an audience of approximately 900 attendees and speakers. This report contains a chronological summary of the key discussions and conclusions from almost all of the workshops, tutorials, conference sessions, and special interest groups.

**A CD-ROM including all of the conference presentations is available for purchase.**

To order the CD-ROM, or to receive more information about this conference, and related future events, go to <http://www.wilshireconferences.com>

### Reproduction Policy

This document is © Copyright 2004 Wilshire Conferences, Inc. It may be copied, linked to, quoted and/or redistributed without fee or royalty only if all copies and excerpts include attribution to Wilshire Conferences and the appropriate speaker(s). Any questions regarding reproduction or distribution may be addressed to [info@wilshireconferences.com](mailto:info@wilshireconferences.com).

**WORKSHOPS**  
**Sunday, May 2**  
**3:30 pm – 6:45 pm**

**Data Modeling Basics**  
**Everything you need to know to get started**

*Marcie Barkin Goodwin*  
*President & CEO*  
*Axis Software Designs, Inc.*

This workshop for novices covered the basics of data modeling, including an explanation of:

- IDEF & IE Methods
- Entities, Attributes & Relationships
- Keys, Cardinality, Recursion & Generalization Hierarchies
- Normalization
- Standards & Procedures

**Abstraction: the Data Modeler's Crystal Ball**

*Steve Hoberman*  
*Global Reference Data Expert*  
*Mars, Inc.*

Steve explained abstraction as redefining entities, relationships, and data elements into more generic terms. He went through an example illustrating how to abstract, and stressed the importance of apply the Abstraction Safety Guide. This guide contains 3 questions we must ask ourselves before abstracting. The 3 questions are "Do we have something in common in our model?", "Is there value in abstracting?", and "Is it worth the effort?" Steve then put abstraction in context with normalization and denormalization. He explained subtyping and meta data entities, and shared with the group a sad story and a happy story where abstraction was used in the production data warehouse. The group then broke down into smaller groups and did a short workshop on abstracting contact information. He ended the session discussing reusable abstraction components and explained universal models and their uses.

**Enterprise Architecture Principles and Values - Straight from the Source!**

*John Zachman*  
*President*  
*Zachman International*

In the Industrial Age, it was the products (airplanes, buildings, automobiles, computers, etc.) that were increasing in complexity and changing. We had to learn how to describe them (architecture) in order to create them and maintain them (that is, change them) over time. In the Information Age, it is the Enterprise that is increasing in complexity and changing. By the same token, we will have to learn how to describe them (architecture) in order to create them to accommodate dramatic increases in complexity and keep them relevant as they change over time. The Framework for Enterprise Architecture defines the set of descriptive representations

that constitutes Enterprise Architecture and establishes a semantic structure to understand the physics of Enterprise engineering.

For those who understood the value proposition, Enterprise Architecture has always been important. Yet it is only relatively recently that the concepts and benefits of Enterprise Architecture have started to be embraced by a significant number of organizations at a senior level.

John explained how and why Enterprise Architecture provides value, and the four reasons why you "do" Architecture including alignment, integration, change management and reduced time to market. Without Architecture, there is no way you can do any of these things.

## **Process Orientation for Data Management Professionals: Proven Techniques for Achieving Support and Relevance**

*Alec Sharp  
Consultant  
Clariteq Systems Consulting Ltd.*

*Business Process* has returned as a hot topic:

- economic and regulatory reality (e.g., Sarbanes-Oxley) demands it
- systems are far successful when implementation is "process-aware"

Adopting process orientation is a great way for Data Management professionals to add value while promoting the importance of their discipline. Some specifics covered:

A common mistake is to identify a functional area's contribution as a complete process, leading to the common problem of "local optimization yielding global suboptimization."

Workflow modeling is a physical modeling technique, which causes difficulty for many analysts who are used to essential (or logical) modeling

Showing the often-ugly reality of as-is processes raises awareness of the true cost of inconsistent, disintegrated data.

Processes only perform well when all of the six enablers work well – workflow design, IT, motivation and measurement, policies and rules, HR, and facilities.

## **A Data Stewardship "How-To" Workshop: De facto, Discipline & Database**

*Robert Seiner  
President & Principal/Publisher  
KIK Consulting & Educ Services/TDAN.com*

New sets of rules, regulations and legislation are beginning to require formalized accountability for data assets and how they are managed across the enterprise. The business and technical value of data stewardship is viewed as insurance, driven by financial reporting, data integration requirements, accounting needs and the audit-ability of data and data focused processes.

Students of the workshop learned about the proven components of effective Data Stewardship programs and how to immediately customize a strategy that will deliver results with limited financial investment. Mr. Seiner focused on best practices and "how-to" develop a pragmatic

and practical solution to implement a data stewardship program. Attendees of this class learned:

- How to formalize accountability for the management of enterprise data
- How to recognize the need to implement a data stewardship program
- How to frame and sell data stewardship in the corporate context
- How to insert data stewardship into data management policy
- How to define objectives, goals, measures/metrics for stewardship programs
- How to align steward roles with data through the “de facto” approach
- How to identify, implement, enforce the “discipline” through process integration
- How to build & implement a stewardship repository or “database”

## **Fundamentals of Data Quality Analysis**

*Michael Scofield  
Assistant Professor  
Loma Linda University*

You can't assess the quality of data without understanding its meaning. And the meaning of data is derived, in part, by its position in a data architecture (both physical and logical). Therefore, Mr. Scofield began with a review of the nature of logical business data architecture, and how it affects the meaning of data. While data architecture is abstract, data model is a way of describing it, and achieving consensus on current and future reality for an enterprise.

Next, we dealt with doing a data inventory. Gathering knowledge about the data asset is both technical and political. One often encounters a lead time or critical path in getting permission from some parts of large organizations to get read access to the data. Then, building up the meta-data, one surveys databases (even topically-organized clumps of data not necessarily in a relational DBMS). Inventory the databases, the tables in each (with their meaning and behavior), and the fields in each table (with meaning, behavior, and quality).

The center part of the workshop focused upon data profiling or understanding the behavior and quality of the data asset. While there are some commercial tools available to support this effort, the crucial element is a cynical data analyst who understands the business. Central to data profiling is the domain study which surveys data behavior in an individual column, showing the anomalous values observed. Once anomalies are discovered, their causes must be determined, often through dialogue with application experts or business experts. In such a dialogue, having a formatted dump of the records in question is useful. The answers should be clearly documented. Graphic techniques were also shown for finding anomalies in the data and understanding them.

Data quality has many facets, including presence, validity, reasonableness, consistency, accuracy, and precision. The method of delivery (ease of use, timeliness, etc.) should not be confused with data quality (even though it often is). So the utility of data (suitability for a particular purpose and availability) is independent to the intrinsic quality of the data.

Finally establishing a data quality program cannot be done by one person. It requires involving the “owners” of all the processes around an enterprise. Data quality measurement can be done by an expert, although it is valuable to provide process owners the means to measure their own data quality on a recurring basis. Only they can improve their processes and correct the systemic factors (often in manual procedures, screen design, but also in business processes and incentives for the data gatherer) which cause the bad data in the first place.

## **New Trends in Metadata: Metadata is not Just Repositories Anymore**

*John Friedrich  
President  
My Krisalis, Inc.*

This workshop explained many of the new concepts that surround meta-data movement and integration. Those who attended now have an understanding and appreciation for the following:

Meta-Data is NOT synonymous with "repository"

The industry, including repository, ETL, BI and DW vendors, are beginning to offer meta-data exchange capabilities

Standards, such as CWM, UML and IDEF, have lead to a proliferation of integration, or "glue", providers that can be leveraged for metadata movement

New meta-data movement issues to contend with include:

- Meta-data warehousing
- Enterprise meta-data application integration
- Meta-data ETL
- Business meta-data intelligence

## **Acquiring, Storing and USING Your Meta Data**

*Doug Stacey  
Area Leader, Allstate Data Management  
Allstate Insurance Company*

Effective meta data management is much more than just a matter of identification and collection. The whole point is that people in the organization...users, analysts and developers...actually USE the meta data you're making available to them. This workshop provided a case study of how they make meta data work at Allstate, including:

Overview of the meta data environment at Allstate

You must put meta data to work!

Capturing meta data

Storage of meta data

The repository

Concept of domains

Meta data for the warehouse environment

Meta data to support run-time applications

Return on investment - thoughts/examples for calculating ROI

## **Unstructured Data Management**

*David Raab  
Partner  
Raab Associates*

Unstructured data management covers text, sound, images, HTML files, and other formats. Managing such data is based on applying category tags that can be sorted, selected and

otherwise managed like structured data. Categories are called "taxonomies"; creating and maintaining a relevant taxonomy often takes more work than expected. Tags can be applied manually, but that's very expensive. Automated systems fall into two broad categories, statistical and semantic. Statistical systems analyze samples of records in each category to identify the words and patterns that distinguish them. They are relatively easy to deploy and computationally efficient, but not necessarily precise. Semantic systems analyze the actual language of a document using preexisting knowledge such as grammatical rules, dictionaries and "ontologies" (descriptions of relationships among concepts). They are better at extracting specific information from documents, as opposed to simply identifying the category a document belongs to. As with any system, defining specific requirements is essential to selecting the right tool for a particular situation.

## **Models, Models Everywhere, Nor any Time to Think A Fundamental Framework for Evaluating Data Management Technology and Practice**

*Fabian Pascal  
Analyst, Editor & Publisher  
Database Debunkings*

The speaker maintains that the majority of data management practitioners operate in "cookbook", product-specific mode, without really knowing and understanding the fundamental concepts and methods underlying their practice. For example: what data means, what is a data model, data independence, etc. This workshop provided a fundamentally correct way to evaluate data management technologies, products and practices. It helped practitioners understand data fundamentals that are either ignored or distorted in the industry, how to apply these fundamentals in day to day business, and how to use them in the evaluation of the technologies being promoted.

**NIGHT SCHOOL  
Sunday, May 2  
7:00 pm – 8:00 pm**

**Information Assurance**

*Ronald Borland  
Data Architect  
Nationwide Mutual Insurance*

Ron Borland's presentation on Information Assurance addressed the need for enterprises to develop a coordinated approach to improving information quality. Ron's premise is that, only by actively involving all the stakeholders directly in the data quality process can we have a reasonable expectation that they will agree on the success criteria by which the effort is measured. The three key takeaways from Ron's presentation are:

- Get the right people involved early and keep them actively and enthusiastically involved
- Develop a culture that supports a senior escalation point for data quality issues from which there is no appeal
- Insure that every point where the process touches reality adds value and eliminates bottlenecks

**Data Management Practice Maturity Survey**

*Peter Aiken  
Founding Director  
Institute for Data Research*

*Burton Parker  
Principal  
Paladin Integration Engineering*

Over the past two years, the Institute for Data Research has surveyed more than 40 organizations of differing sizes - from both government and industry. The results of this survey are permitting the development of a model that can help organizations assess their organizational data management practices. Good data management practices can help organizations save the 20 - 40 % of their technology budget that is spent on non-programmatic data integration and manipulation (Zachman). This talk described the Data Management Practice Maturity Survey and presented the results to date.

**Introduction to ORM**

*Terry Halpin  
Professor and VP (Conceptual Modeling)  
Northface University*

Effective business rules management, data warehousing, enterprise modeling, and re-engineering all depend on the quality of the underlying data model. To properly exploit database technology, a clear understanding is needed as to how to create conceptual business models, transform them to logical database models for implementation on the chosen platform, and query the populated models. Object-Role Modeling (ORM) provides a truly conceptual way to accomplish these tasks, facilitating communication between the modeler and the domain expert.

## **Lies, Damn Lies, and Enterprise Integration: A Fair and Balanced Perspective**

*Evan Levy  
Senior Partner  
Baseline Consulting*

Evan Levy gave attendees a lively look at data integration alternatives. Various examining Enterprise Resource Planning (ERP), Enterprise Application Integration (EAI), Enterprise Information Integration (EII), and data warehousing, Evan analyzed the comparative applications of each of the four environments, deconstructing their functional differences. He provided a perspective on the vendors who play in each of the four arenas, and discussed the relative strengths and weaknesses of the various integration approaches, focusing--often with his tongue firmly planted in his cheek--on specific business drivers best-suited for each. Evan relayed some real-life client experiences with different integration technologies, and concluded his talk with predictions for the future of all four integration options.

## **Metadata and Grid Computing**

*Frank Kowalkowski  
President  
Knowledge Consultants, Inc.*

*Clyde Laakso  
President  
Lake, Vesta and Bright*

Grid computing is getting a lot of press today as the 'next big thing' that will impact the enterprise. It is true that grid computing exists today in several variations such as the Seti project and platform load-sharing in an enterprise that offloads processing capabilities across machines and takes advantage of unused capacity to run applications. However, making the concept useful, practical and a 'value add' for running applications on an external grid for the enterprise requires some significant convergence of technologies and disciplines. To make the grid idea useful within an enterprise we need an extended grid concept. Moving the extended grid concept to the business landscape and more specifically to the enterprise environment requires four important capabilities to make the grid viable; clean and standardized metadata and metadata services, a bill of materials for an application, a routing list for execution and a costing mechanism for billing.

## **Semantics in Business Systems**

*Dave McComb  
President  
Semantic Arts*

Dave McComb introduced semantics as the underlying discipline behind much of what we do as data modelers. After outlining where Business Vocabulary, Taxonomy and Ontology fit in, he introduced several new concepts, including: Fork Genera (why it is that most of the knowledge we have is at the middle level of a taxonomy), and how the Semantic Web is structured. He then described a framework for reasoning about the semantic content of our systems, which included factoring meaning into semantic primes, contexts and categories.

**TUTORIALS**  
**Monday, May 3**  
**8:30 am – 4:45 pm**

**Business-Focussed Data Modeling**  
**Unifying Conceptual, Logical, Physical Models**

*Graham Witt*  
*Independent Consultant*

This tutorial started from the position that while there is general understanding about what a physical data model is, there are many different interpretations of the terms “conceptual data model” and “logical data model”. The underlying premise of the tutorial was that the process of developing a database design that accurately reflects business information requirements must start with the production of a model of business concepts that not only uses business terminology but includes structures that reflect reality rather than the dictates of normalisation or the need to fit a pre-SQL3 relational straitjacket. Graham adopted the term “conceptual data model” for such a model, which may include any of the following:

- supertypes and subtypes
- business attribute types, which describe an attribute’s behaviour, rather than DBMS datatypes
- category attributes rather than classification entities
- complex attributes, with internal structure
- multi-valued attributes
- derived attributes & relationships
- attributes of relationships
- n:n relationships
- n-ary relationships.

The tutorial also covered practical techniques for modeling with these component types when using a CASE tool that may not support one or more of them, as well as ways of converting them to a pre-SQL3 relational model (“logical” or “first-cut physical” depending on your preference).

A conceptual data model cannot be developed without a thorough understanding of the business information requirements, so practical tips for requirements discovery were also provided, in particular the development of an Object Class Hierarchy. Rather than being a traditional E-R or Object Class Model, an OCH consists of Business Terms (which may be Entity, Relationship or Attribute Classes or even Instances), each with Definitions and organized hierarchically with inheritance of properties. It provides a high level of business buy-in, and a genuinely reusable resource, since application projects retain confidence in the validity of the content. Some practical tips for business rule discovery and documentation were also imparted, as well as a dissertation on the importance of unambiguous names and definitions.

Review of the conceptual data model is essential so Graham then described how to generate a set of assertions (or verbalizations) about the model which can be reviewed individually, each reviewer disagreeing, agreeing or seeking clarification.

## **Developing Better Requirements and Models Using Business Rules**

*Ronald G. Ross*  
*Principal*  
*Business Rule Solutions, LLC*

*Gladys S.W. Lam*  
*Principal*  
*Business Rule Solutions, LLC*

This tutorial explained how the business rule approach can improve your entire requirements process. Starting with the business model, the speakers identified relevant deliverables and showed where business rules fit in with them. Specifically, they showed how business rules address the issues of motivation and guidance – in other words, the question of “why.” They detailed how you can use business rules to develop business tactics in a deliverable called a Policy Charter.

Continuing, they focussed on the system model and how business rules fit with each deliverable. Special emphasis was given to how business-perspective rules differ from system-perspective rules, and what you need to do to translate between the two. Practical refinements to system model deliverables were examined, not only to exploit business rule ideas, but also to maintain a consistent focus on validation and communication from the business perspective.

Finally, Mr. Ross and Ms. Lam offered guidelines for how business rules should be expressed, and how they can be managed more effectively.

## **Using Information Management to Sustain Data Warehouse**

*John Ladley*  
*President*  
*KI Solutions*

Information management is vital to sustaining a data warehouse. This means the DA areas are now accountable for more than the models. John Ladley reviewed how DA areas are on the front line of developing business cases, enabling new business processes, and assisting in culture change. Several key techniques were presented that leverage the strengths of DA areas (abstraction , modeling, governance) to helping organization better align and sustain data warehouses.

## **How to Develop an Enterprise Data Strategy**

*Sid Adelman*  
*Principal*  
*Sid Adelman Associates*

### **Current Status in Contemporary Organizations**

Very few organizations, large or small, have a well-defined enterprise data strategy. If asked some will point you to dusty and outdated volumes of data base standards, usually geared specifically to their Relational Data Base Management System (RDBMS). The more advanced

organizations will have a subset of standards and perhaps a documented strategy on portions of what should be included in an overall strategy. In most organizations, the value of data is not well understood. Data is considered the province of the department that creates it and this data is often jealously guarded by that department.

Data is usually addressed on a piecemeal basis. A company will launch an effort to choose its preferred RDBMS or will attack a database performance problem when response time becomes excessive. Rarely do organizations work from the big picture and as a result suboptimize solutions, introduce programs which may have a deleterious effect on the overall enterprise, cause inconsistencies that result in major efforts for interfacing or develop systems that can not be easily integrated

### **Why an enterprise data strategy is needed**

Not having an enterprise data strategy is analogous to a company allowing each department and each person within each department to develop their own chart of accounts. The empowerment would allow each person in the organization to choose their own numbering scheme. Existing charts of accounts would be ignored as each person exercised his or her own creativity. Even to those of us who don't wear green eye shades, the resulting chaos is obvious.

The chaos without an enterprise data strategy is not as obvious, but the indicators abound: dirty data, redundant data, inconsistent data, and users who are becoming increasingly dissatisfied with the performance of IT. Without an enterprise data strategy, the people within the organization have no guidelines for making decisions that are absolutely crucial to the success of the IT organization. In addition, the absence of a strategy gives a blank check to those who want to pursue their own agendas. This includes those who want to try new database management systems, new technologies (often unproved), and new tools. This type of environment provides no checks for those who might be pursuing a strategy that has no hope for success.

An enterprise data strategy should result in the development of systems with less risk and a higher success rate. It should also result in much higher quality systems. An enterprise data strategy provides a CIO with a rationale to counter arguments for immature technology, and data strategies which are inconsistent with existing strategies.

### **Vision**

The vision of an enterprise data strategy that fits your organization has to conform to the overall strategy of IT which in turn must conform to the strategy of the business. The vision should conform to and support where the organization would want to be in five years.

## **Enterprise Messaging Modeling and Data Integration Bringing Data Skills to Service Oriented Architecture**

*Dave McComb  
President  
Semantic Arts*

This was a highly interactive and participatory session, where the audience worked their way through a "makeover" from data modelers to message modelers. The tutorial combined elements of why enterprises are moving to Service Oriented Architectures, and using Messages to achieve integration. Through a series of structured exercises the participants were able to experience the impact of splitting existing systems into functional partitions, and shared services

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

and then re-integrate them with messages. The session also contained a methodology for doing this at an enterprise level and guidelines for achieving loose coupling between applications and services.

## **Enterprise Metadata: The Art and Science of Best Practices**

*R. Todd Stephens  
Director  
BellSouth*

Enterprise metadata is moving to the forefront of most major organizations and we must ensure that we succeed at implementation and integration through out the organization. Todd discussed and provided a five year learning curve for the attendees of this tutorial. The session reviewed a metadata framework that covered everything from the raw material of metadata, products, processes, customer service and the metadata experience. Other items covered included:

- Enterprise Metadata Environment
- The Architecture of an Enterprise Metadata effort
- The Project and Implementation side of an enterprise effort
- The importance of Usability in Metadata
- Technical Architecture of the Repository Collection
- The Principle of Success around the service side of delivery
- Key Lessons Learned

Metadata and the principles that define this technology must be expanded into the other areas of the enterprise environment. While data was the key interest at this conference, we cannot ignore the entire collection of assets in the corporation.

## **Building the Managed Meta Data Environment**

*David Marco  
President  
Enterprise Warehousing Solutions, Inc.*

The tutorial "Building the Managed Meta Data Environment (MME)" covered a great many topics. During the first quarter of class the fundamental of meta data were discussed. Including technical meta data, business meta data, and a very detailed discussion on the return on investment that a MME provides. Over a dozen MME use examples were presented to the audience.

The second and third quarters of the tutorial were focused on the managed meta data environment and its 6 key components: Meta Data Sourcing Layer, Meta Data Integration Layer, Meta Data Repository, Meta Data Management Layer, Meta Data Marts, and the Meta Data Delivery Layer. Each of these six components were discussed in great detail. In addition, several implementation best practices were illustrated. The last quarter of the seminar focused on the various tool vendors in the meta data management space. Six vendor tools were discussed in detail and all of the meta data tool questions from the audience were fielded. The tutorial ended with two real-world MME case studies.

## **XML Database and Metadata Strategies**

*James Bean*  
*Chairman and Co-founder*  
*Global Web Architecture Group*

Determining the most effective strategy for persisting (storing) XML in a relational or object-relational database requires careful research and evaluation. A "best practices" approach incorporates activities and techniques that are analogous to those used for more traditional data architecture efforts.

Business and data requirements, constraints, and capabilities can be used to guide selection of the best XML database storage strategy. The most common of which include:

- Store the entire XML document as a complete entity
- Decompose (e.g. "shred") the XML document and store separate data values in tables and columns
- Store a combination of XML document fragments and separate data values
- Store the XML document external to the database, but link if possible

Further, there are a number of complexities and challenges to also be addressed (i.e. character encoding, data type disparity, schema type support, etc.)

## **Mastering Reference Data**

*Malcolm Chisholm*  
*President*  
*Askget.com Inc*

This presentation underscored reference data as a special class of data in an enterprise information architecture – a class that has its own special behavior and its own special management needs. Data Administrators have to understand what these needs are and create an organization and tool set that enables reference data to be managed effectively. Failure to do so leads to data quality problems, severe limitations on the ability to share data, and the incorrect implementation of business rules. Practical advice was given on how to achieve good reference data management, and the session was marked by a lively interaction that demonstrated the relevance of this topic to the attendees.

## **Data and Databases**

*Joe Celko*  
*Professor and VP of Relational Database Management Systems*  
*Northface University*

This technical class examine scales, measurements and how to encode data for a database. Part two of the class examined how to apply the theory and techniques of part one in an SQL database environment.

**NIGHT SCHOOL  
Monday, May 3  
5:00 pm – 6:00 pm**

**One Purdue Information Factory**

*Brad Skiles  
Director, Data Services and Administration  
Purdue University*

Purdue is preparing to migrate from diverse legacy applications, some of which are over 30 years old, to an ERP environment. Prior to this move, Purdue will be implementing a new data architecture called the One Purdue Information Factory (OPIF). This presentation presented how the OPIF attempts to imbue quality in Purdue data through the practical application of these three components. Specifically, Purdue's plan to:

- implement a scalable and integrated architecture based upon the Bill Inmon information factory architecture
- capitalize on a robust metadata engine
- assess and correct data content in the legacy systems
- implement a state-of-the-art business intelligence tool suite

**Bare-handed Data Modeling: A Polemic**

*John Schley  
Senior Data Administrator  
Principal Residential Mortgage, Inc.*

While we data people feel that our role and deliverables are very valuable for the companies we work for, there are a lot of people on the business and technical side that do not agree. They are not all idiots.

This presenter began by listing the three main deliverables of a data modeling effort -- data models, business definitions, and DDL. He detailed how each can be nearly worthless to our business and technical partners.

Then the presenter listed 19 ways to refocus on the real value of data modeling--not as a competency in itself but as a way to introduce the field of data management to the enterprise. In short, we need to more fully support the business by reaching up and across the Zachman Framework to address real issues.

**What is Why? What is How? And What is What?  
Distinguishing Business Rules From Data And Processes**

*Tom Yanchek  
Manager-Corporate Repository/Architecture  
United Parcel Service*

Distinguishing a business rule from the data it supports or a process it enforces can be tricky and often times confusing. This presentation discussed:

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

- How to overcome common roadblocks when identifying business rules
- How to separate data, process and rules without losing their dependencies
- How the importance of employing an organized Business Rule Strategy helps build better enterprises
- How the importance of employing an organized Business Rule Strategy helps enhance or possibly consolidate existing portions of an enterprise
- Tips and Techniques for identifying a business process, a business rule or business data
- How to measure the success of the approach and how to identify, employ and implement quality improvement procedures
- How to minimize or significantly reduce development costs and maintenance costs – in some cases 25-35%
- How to maintain business processes, a business rules and business data for verification, acceptance and reasonability

**Meta Data Starts with an Enterprise Architecture Data Model**

*John Sharp  
Principal Consultant  
Sharp Informatics*

Collecting, analyzing and reviewing your organization's information status and future goals are required for successfully implementing an Enterprise Architecture. This presentation established the rules to make your Enterprise Architecture project successful. The basic model contains major objects such as principles, technology, standards, business drivers, processes and applications. The relations between these objects can be expressed as simple sentences. A simple procedure was shown for extending the EA Data Model to include additional knowledge by converting simple true statements into valid fact types. Examples of improvements in the quality and quantity of collected data using the application created from an EA Data Model were presented.

**Data Management Using Metadata Registries**

*Judith Newton  
Computer Specialist  
NIST*

The adoption of XML as the information interchange format for the Web presents a set of challenges and opportunities for data managers. While XML makes it easy to describe the format of information objects and the relationships among them, it has limited provisions for conveying the semantic content of the objects. Metadata registries based on the ISO/IEC standard 11179 supplement XML schema descriptions and all other data applications with descriptive and relational metadata, allowing all representations of information to be mapped together in a single resource. The new version of 11179 introduces a Metamodel that allows incorporation of conceptual and logical level objects along with multiple representations of these objects. This makes it a vessel uniquely suited to store metadata for multiple applications. Principles developed for the establishment of standardized names through naming conventions can be used to administer namespaces and the objects contained in namespaces.

## **Integrating Unstructured Data**

*Michael Reed  
Principal Consultant  
Enterprise Warehousing Solutions, Inc.*

This night school class covered the concepts of unstructured data and the issues surrounding integration with more structured (e.g. database) data. Entity Extraction tools as a means of extracting tagged information from documents and merging it in existing structured tables was discussed. The class also concentrated on the application of accepted industry standards and protocols to unstructured metadata. Standards covered included Dublin Core, ISO 11179, the Resource Description Framework (RDF), and the Java Metadata Interface (JMI).

Also discussed were concepts around classifications, including taxonomies and ontologies. A poll of the class showed that, although most firms were interested in developing classification schemes, almost none had yet begun to do so. The summary discussion emphasized the point that in order to get a true view of "all that we know", both structured and unstructured data must be collated, classified, and made available to end-users.

## **Conducting Database Design Project Meetings**

*Gordon Everest  
Professor  
University of Minnesota*

Dr. Everest provided a synthesis of his experiences in gathering database design requirements through interviews and conducting database design project meetings. Reflecting on the lessons learned, he offered a set of ideas and best practices which should be considered by anyone attempting to lead a database design project, including.

- Interviews vs. facilitated group sessions
- Picking the users to interview or invite to the table.
- Preparations before the interviews or before the meetings
- Accelerated (one meeting possibly over several days, e.g., JAD) vs. extended series of meetings
- Findings of an experiment which compared the two approaches to group sessions

## **To Laugh or Cry 2004: Further Fallacies in Data Management**

*Fabian Pascal  
Analyst, Editor & Publisher  
Database Debunkings*

A lot of what is being said, written, or done in the information management field by vendors, the trade press and "experts" is increasingly confused, irrelevant, misleading, or outright wrong. The problems are so acute that, claims to the contrary notwithstanding, knowledge, practices and technology are actually regressing! This presentation exposed some of the persistent misconceptions prevalent in the information/data management field and their costly practical consequences.

**SPECIAL INTEREST GROUPS**

**Tuesday, May 4  
7:15 am – 8:15 am**

**Automating Conversions Between Metadata Standards Using XML and XSLT**

*Kara Nance  
Professor of Computer Science  
University of Alaska Fairbanks*

*Brian Hay  
Computer Science  
University of Alaska Fairbanks*

Many data system designers and organizations are faced with a dilemma when deciding which metadata standard(s) to apply when designing data systems. In some cases, datasets don't easily fit into the scope of current defined domains, but might in the future as standards for new domains are developed. The increased emphasis on new metadata standards coupled with the user demand for seamless data portals, creates conflict for many data system designers, often delaying production by lengthening the design phase of the system development life cycle. This session presented a process through which the conflict can be mediated and design delays minimized. XSLT and XML can be used to automatically transform datasets between metadata standards to create a single data format which can then be manipulated as a single dataset.

**KEYNOTE PRESENTATION**

**Tuesday, May 4  
8:45 am – 9:45 am**

**Database Graffiti: Scribbles from the Askew Wall**

*Chris Date  
Independent Author, Lecturer, Researcher and Consultant*

This presentation was based in part Chris Date's writings and columns over many years. It consisted of a series of quotations, aphorisms, and anecdotes - seasoned with a fair degree of personal commentary - that were directly or indirectly relevant to the subject of database management. The presentation was not technically deep, but several serious messages lay just beneath the surface. Topics included:

- Objects and objections
- Normalization, networks and nulls
- The role of simplicity
- The joy of self-reference
- Relational misconceptions
- Some good quotes
- Books and book reviews

## **CONFERENCE SESSIONS**

**Tuesday, April 29  
10:15 am – 11:15 am**

### **Entropy, Energy, and Data Resource Quality**

*Michael Brackett  
Consulting Data Architect  
Data Resource Design & Remodeling*

Data quality is becoming increasingly important for every public and private sector organization, particularly as organizations are being held accountable for the quality of their data. However, there are many different perspectives of data quality ranging from developing databases to integrating data to business information. There are also many perceptions about what quality is and the cost of Quality. This presentation steps back and takes a basic look at quality from the standpoint of physics - entropy and energy. It covers basic topics about:

- How data go bad
- What it takes to maintain data good
- The costs for collecting good data
- The costs of integrating disparate data
- Is data quality really free
- When is quality good enough

### **Adding Value in a Changing Environment - A Case Study**

*Duncan McMillan  
Enterprise Architect (Information)  
Westpac Banking Corporation (NZ)*

*Paul Brady  
Project Data Architect  
Westpac Banking Corporation (NZ)*

The presentation was based on experiences covering the last 15 years. It was found that to successfully initiate and sustain IM and Architecture approaches in a changing environment within Westpac, it had been important to:

- Be aware of the internal and external environment
- Realise that change is inevitable. Understand the nature of change. Adapt to change, and treat changes as opportunities to add value.
- Understand that value is about delivering and also about perception.
- Use an approach that ensured visibility and included directly dealing to organisation-critical changes, and also to culture, progression of the underlying foundation and buy-in.
- Use a portfolio-based approach to reduce risk
- Have a foundation that evolved
- Understand that a critical part of the foundation is the organisation's relationship capability
- Fund relationship training for all staff, and encourage opportunities for relationship and network building.

From this came 4 key messages:

- **Change is an opportunity**
- **Add value in a visible way**
- **Evolve a foundation**
- **Don't forget the importance of relationships.**

## **10 Great Things Which Data Administrators Can Do with Use Cases and Class Diagrams**

*Susan Burk*  
*Director - Systems Architecture*  
*MassMutual Financial Group*

This discussion examined how Data Management can work with use case models and class diagrams to support data definition and logical data modeling, as well as ways for DAs to use their own strengths to support better understanding of business requirements.

- #10 Gain Context for operational systems
- #9 Define terms
- # 8 Find entities and rules
- #7 Balance Data and Behavior
- #6 Encourage normalization of process
- #5 Find data for rules
- #4 Gather volumetrics
- #3 Use the class model to determine or map to an existing data model.
- #2 Use these techniques to re-engineer DA
- #1 Get involved earlier in projects.

## **A Practical Guide for Inventorying Your Data Assets**

*Perry Benjamin*  
*Data Management Specialist*  
*The Boeing Company*

An inventory of data resources can provide the metadata visibility necessary for an enterprise to more easily integrate its computer applications and respond more quickly to business and technological change. The approach taken to develop and populate the data inventory in use at the Boeing Integrated Defense Systems - Long Beach site was presented, including:

- Scope and organization of the metadata
- Construction of an enterprise common business object model
- Metadata capture and change management
- Components of the inventory system



## **Are Business Rules Meta-Data In Disguise?**

*Terry Moriarty  
President  
Inastrol*

During this presentation, the following topics were discussed:

**\* What are business rules?**

The separation of business concepts, terms and business rules was provided, with examples of how different communities use different terms to refer to the same business concept. Within an enterprise, each community wants to state its business rules in its own vocabulary, but still be able to integrate with all the other communities. A business-oriented business rule classification scheme was provided, using a case study of buying a house within various golf communities.

The classification scheme is:

- Qualification Rules: Do I qualify for this business state? Who is in this business state?
- Validation Rules: Am I in a Valid State?
- Configuration/Value Setting Rules: Put me in a valid business state.
- Price Me. Compute me. Configure me.
- Process Rules: What do I do next?

**\* What is Meta-data?**

Metadata was discussed using the Data->Information->Knowledge->Wisdom hierarchy. What you consider to be metadata varies based on which perspective you are at. Business Rules fall into the knowledge level where you internalize the rules and just "know" how something is supposed to work. Business rules are part of business data, therefore requiring that metadata achieves the status of business data.

**\* Aren't Business Concept/Fact models really data models?**

To answer this question, the different types of data models were explored: conceptual, logical and physical. If you maintain a model from which the business can be read, it is probably the same as a business concept model. However, if your data model has the characteristics of a dynamic model which allows new data to be supported very rapidly into the information resource, then you probably need a business concept model also. The Zachman Enterprise Systems Architecture framework was used to describe a dynamic model and techniques for mapping an organization's business concepts into a dynamic model's framework was provided.

## **Building An XML-Based Data Standard**

*Bill Hopton  
Systems Consultant  
Manulife Financial*

If the spaghetti of point-to-point interfaces is the wrong way to integrate multiple systems, re-creating the spaghetti inside an EAI tool is not a big improvement. What is needed is a catalogue of standard messages to de-couple the sources from the targets. In our attempt to create such a catalogue (expressed in XML Schema) we had success enriching the messages inside the middleware using a cross-reference database. We also found that some data transformations were so complex that we were unable to achieve 100% de-coupling with the

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

budget we had. However, by making this attempt we feel we have a good start that we can build on in the future, which we hope will involve the ISO15022 financial services standard.

**From Afterthought to Asset: The Business Value of Data**

*Jill Dyche  
Vice President and Partner  
Baseline Consulting Group*

Citing recent research claiming that business data is growing exponentially, author and consultant Jill Dyche discussed why senior business executives are (finally!) turning their heads in the direction of information as a competitive weapon. Dyche explained why yesterday's practice of "data as enabler" is now ceding to a new and innovative business Zeitgeist of "data as corporate asset." She explained the impact this phenomenon has had on businesses, as well as on development teams. Dyche introduced the concept of the Information Center of Excellence as representative of data's evolution, and discussed the role of the ICE as the company's fulcrum for data delivery.

**CONFERENCE SESSIONS**

**Tuesday, May 4  
11:30 am – 12:30 am**

**Achieving Data Excellence: Roadmap to Success**

*Larry Dziejczak  
Technology Manager  
Johnson & Johnson*

*Maggie O'Hara  
Assistant Professor of MIS  
East Carolina University*

Data quality, or the lack thereof, has become a hot topic in business. This is not surprising, given that in 2001, The Data Warehouse Institute estimated that dirty data cost US businesses \$600 billion. While much attention has been given to defining data quality and to the impact of poor data on the business, less attention has been given to the specific processes involved in improving data quality. Data Quality Improvement (DQI) to achieve data excellence was the focus of this presentation. Using a data life-cycle approach, the presenters demonstrated clear, concrete steps organizations can take to improve data quality at all life-cycle stages: capture, storage, and access.

**The Training and Mentoring of Data Analysts**

*Janet Siebert  
Data Architect  
Lexonics*

Three main theories from educational psychology converge to provide a framework for the training and mentoring of knowledge workers in general, and data analysts in particular. The constructs are Benjamin Bloom's Taxonomy (1956), Howard Gardner's Multiple Intelligences (1983), and Mihaly Csikszentmihalyi's Flow (1990). Bloom argued that thinking skills can be classified into a hierarchy of increasing complexity. Levels of the hierarchy are knowledge, comprehension, application, analysis, synthesis, and evaluation. Gardner argued that there are a variety of different types of intelligence, including linguistic, logical-mathematical, musical, bodily-kinesthetic, spatial, interpersonal, and intrapersonal. Csikszentmihalyi suggests that we are happiest as human beings when we are engaged in challenges well matched to our skills. Successful data analysts leverage multiple intelligences (linguistic, logical-mathematical, spatial, and interpersonal) and operate at relatively high levels in Bloom's taxonomy. Challenging data analysts requires an awareness of each individual's intelligences and thinking skills, and the crafting of activities that are appropriate to that individual.

## **Verbalizing Business Rules**

*Terry Halpin*  
*Professor and VP (Conceptual Modeling)*  
*Northface University*

Although business rules may be implemented in many ways, they should first be specified at the conceptual level, using concepts and language easily understood by the business domain experts who are required to validate the rules. Business rules come in many varieties, such as constraints and derivation rules, and may be specified using graphical and/or textual languages. The focus of this presentation was on expressing a wide variety of business rules in a high-level, textual language that has been designed for expressibility, clarity, formality, generality, and localizability.

- Using positive, negative, and neutral verbalization patterns for rule validation
- Relational-style verbalization vs Attribute-style verbalization
- The use of context for verbalization of complex rules and rule sets
- Verbalizing rules involving projections over multiple associations of any arity
- Verbalizing reaction rules
- Making verbalizations formal and executable

## **Metadata Management Converges with Business Modeling**

*Joshua Fox*  
*Sr. Software Architect*  
*Unicorn Solutions Inc.*

Metadata repositories help you understand your metadata, and business models help you understand your business. However, metadata repository projects often fail, when the metadata's meaning is ambiguous and unclear; modeling projects often fail when developers ignore the model.

The two gain value when they are integrated in a semantic metadata repository. Metadata gains semantics, and models gain a firm grounding in the actual enterprise infrastructure. This results in improved quality and agility in the management and development of information assets. Existing assets can be reused, and redundant assets can be eliminated. The enterprise then unites on a shared understanding of business concepts and processes.

## **Successful Meta Data Management in Complex Siloed Systems**

*Sean Goggins  
Principal Consultant  
Ciber, Inc.*

- Manufacturing, medical research and the emerging “informatics” labeled disciplines are most often substantially concerned with the existence of inconsistently labeled and classified data across 20 or more small “systems” that compose an organization’s complete super-system for data creation and analysis in these areas. It is also common for disparate versions of systems like this to be deployed in 10 or more specific instances across such organizations or, in some cases, customers. Traditional approaches to data management and meta data management provide a strong but incomplete foundation for successfully adding value to systems of this scale. Participants in this session gained a comprehensive, “Big Picture” view of the challenges to consider in the development of a metadata management strategy for organizations with complex systems that support real time manufacturing, clinical research and informatics oriented domains.

## **XML Tools Comparison: Cutting through the hype - what works for Data Management?**

*Peter Aiken  
Founding Director  
Data Blueprint*

*David Allen  
Chief Executive Officer  
Data Blueprint*

Much of the hype generated around XML comes from and revolves around the various XML tools and XML features of existing tools. This session demonstrated the functionality of various classes of XML tools including:

- XML editors
- Translators
- XML parsing technologies
- Servers

## **Database Trends: The DBMS Landscape Circa 2004**

*Craig Mullins  
Director, Technology Planning  
BMC Software*

Craig Mullins outlined several of the macro trends that are impacting database management and database administration. Chief among these trends were short release cycles of DBMS versions, rising complexity driven by DBMS technology subsuming technology that was previously free-standing, non-stop data volume growth, and the need for non-stop availability

driven by Internet-enabling database applications. He spoke about the need to be able to prioritize DBA tasks based on knowledge of the business value of the data being impacted. He proposed the need for intelligent automation through heterogenous database administration using a central console to mask the differences between databases as a solution to the difficulty of managing database systems.

## **CONFERENCE SESSIONS**

**Tuesday, May 4**

**1:45 pm – 2:45 pm**

### **Using an Enterprise Portal to Drive Global Data Architecture at Citigroup**

*Denis Kosar  
VP Data Architecture  
Citigroup*

This presentation provided the audience with an overview as to how Citigroup's Global Data Architecture Group is dealing with the new pressures that the regulators, business drivers, and the economy have imposed on the way we bring ROI back to the business. In other words, "A survival manual for Global Data Architecture in troubled times". It details examples as to how an enterprise portal can greatly enhance Data Architecture's ability to perform and communicate its role across the enterprise. The key to its strategy is MENTOR, SHARE and EMPOWER. Providing an advisory role, special services, and a "re-use" strategy are key components for measuring the effectiveness through monthly metrics that can be reported.

This presentation seemed to be right on target judging by some of the questions asked, post presentation dialog, and references made in other presentations. In today's troubling times we are being asked to do much more with less and I would hope that this presentation has provided the data professional with a game plan and strategy to not only survive but to make a difference to the enterprise.

### **International Cross-Enterprise Shared Data A Framework for Business Implementation**

*Patrick Curry  
UKCeB*

The goal of securely sharing sensitive data across enterprises and across international boundaries is fraught with security, commercial and regulatory pitfalls. This presentation described the significant progress being made by Defence primes with transatlantic governments support to enable large Defence contracts to benefit from real e-business advantages whilst ensuring compliance of different nations respective national security, data protection, privacy, export control, Sarbanes-Oxley regulations. The description of the Framework and associated guidance highlighted developments in:

- identity management (of personal and assets)
- secure cross enterprise logical information architecture,
- multi-national export control compliance business rules,
- applying DRM

## **Enterprise Data and Process Modeling for the Telecommunications Industry - A case study using the E-TOM model**

*Robert Mattison*

*CEO*

*Excellence In Telecommunication*

In this presentation, Rob Mattison shared his experience in the deployment of an enterprise data model for a start up telco in Asia. By making use of the telecom industry standard E-TOM (Telecommunications Operations Model), and gaining top management commitment to doing it "right the first time", the organization was able to deploy a highly effective and flexible data model (and supporting database structure) that is helping them to attain and maintain market leadership in all operational areas. E-TOM was utilized to create an entire enterprise reporting, data warehouse, data mining and operational process quality infrastructure.

## **The Role of Metadata Management in the Construction of a Successful Data Warehouse**

*James Montague*

*IT Manager-Data Architect-Enterprise Information Management*

*3M Company*

3M's large and diverse Data Warehouse requires a robust, active metadata (MD) repository. It's critical to its success to store MD, supporting the entire construction process: data design, analysis, development, and testing & quality assurance. 3M's MD solution is a best practice in many ways:

- Manages data legacy from source system through each transformation.
- Captures subject areas data structures.
- Audits the DW, detecting undocumented changes.
- Provides web reports to DW developers and business users.
- Generates data models.
- Manages 'canonical formats' for internal ETL.
- Manages business rules and error conditions.
- Defines tolerance: data completeness and accuracy.

## **Creative Metadata on a Shoestring: Inexpensive Metadata Tools**

*Bonnie O'Neil*

*President*

*Westridge Consulting*

This presentation provided an overview of several very helpful, inexpensive tools:

- Infoselect: (price range: \$50 to \$250) [www.miclog.com](http://www.miclog.com)  
Unstructured Data tool: Helps organize and search through very large documents and find a text string
- MindManager: (price range: \$50-\$300) [www.mindjet.com](http://www.mindjet.com)  
Brainstorming software: Helps organize thoughts gathered from meetings; also useful for source to target mappings for data migrations.

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

- KnowledgeDriver: (\$5K) [www.knowledgedriver.com](http://www.knowledgedriver.com)  
Inexpensive data profiling software
- Teamroom: Comes with Lotus Notes; other software can do this too.  
General functionality: Allows you to post documents and allows users to post and subscribe, and create conversation threads. Very useful for obtaining user feedback from geographically distributed users.
- BRSRuleTrack: (\$5K) [www.brsolutions.com/ruletrack.shtml](http://www.brsolutions.com/ruletrack.shtml)  
Metadata Repository for Business Rules.

## **Metadata: How Far Can it Take Us?**

*Doug Lenat  
President  
Cycorp*

After a near 20-year effort to convert human common sense into a form that computers can understand and use, the Cyc Knowledge Base (pronounced “psyche”) has recently emerged into the commercial world. Cyc is like a “super-metadata” engine – providing layer upon layer of additional understanding and context to the specific metadata already in your systems. These extra layers of metadata then open the door for all sorts of new and exciting applications. Cyc is a significant technological achievement that will likely change the metadata landscape forever.

This session described:

- What Cyc is and the CYC Knowledge Base?
- How Cyc provides you with metadata?
- How this is qualitatively different from what metadata usually provides?
- How the layering of more and more metadata helps? What changes occur as we add these additional layers?
- Application areas in data management:
  - Integrating disparate databases
  - Data quality
  - Better search and querying
  - Semantic web
  - Security and compliance

## **Assessing BI Suites and Platforms for Performance Excellence**

*Jonathan Wu  
Senior Principal  
Knightsbridge Solutions*

Business intelligence (BI) provides:

\* The ability to easily access information for analysis, reporting and decision-making purposes without having to understand the associated complexities of the structured data, technical architecture or schemas from transactional systems.

\* Access to information ranging from sophisticated cross-functional enterprise analytics to trend analysis and potentially transactional data

- Market share
- Customer retention
- Operational efficiency

\* Access to information to monitor business activities and performance

BI suites and platform technologies have significantly evolved over the last 5 years. This presentation reviewed the functionality of the leading BI suites and platforms, and addressed the following:

- Leading product vendors
- Technical architectures that provide optimal performance
- Approach to software selection
- Matching needs with the corresponding technology

## **CONFERENCE SESSIONS**

**Tuesday, May 4  
3:15 pm – 4:15 pm**

### **PANEL: Is Data Modeling Alive and Well?**

*David Berger*

*Graham Witt*

*Terry Quatrani*

*Terry Halpin*

*Alec Sharp*

We live in a world of ERP, packages, short-term solutions, tactile projects and outsourcing. Some organizations have replaced corporate data architecture areas with project teams that include database administrators but may not include data modelers. Others only develop physical models and skip development of the logical model.

Nonetheless in many organizations, data modeling is indeed alive and well. Some positive moves in these organizations (that others would do well to emulate) include:

- Moving up the food chain – getting out of tool-driven detail and doing more high-value conceptual modeling
- Adapting to the new reality –reverse-engineering conceptual models out of existing systems to highlight business impact
- Passing our selfish gene to new hosts – helping every developer want to become a competent data modeler
- Acting local while thinking global – giving up on producing the perfect, enterprise-wide model
- Moving beyond being a speed bump – helping out in any possible way as a means of building up credibility

### **Enterprise Common Data Architecture - Roadmap to Integration**

*Daniel Paolini*

*Director, Data Management Services*

*State of NJ Office of Information Technology*

An Enterprise Common Data Architecture (ECDA) is a collection of related tools and technologies, along with standards and policies and the methodology and the expertise to employ them. The architecture enables real-time operational integration as well as the delivery

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

of integrated analytical data to different communities in the format that each requires. The creation of an ECDA is a major yet essential commitment to any long-term strategic initiative to support data reusability. This architecture forms the foundation for collecting, storing, managing and controlling privacy of and access to data on an enterprise basis. The presentation discussed:

- The problems and challenges of a legacy environment
- The opposing needs of operational and analytical systems
- The process to create integrated operational systems'
- The rational creation of analytical solutions
- How to get IT staff, executive management and business users to see the big picture

**Can the Warehousing Center of Excellence implement a Meta Data Repository?**

*Lowell Fryman  
President  
Metrics Consulting Corp*

This session explored the real-world implementation of an enterprise meta data repository by the Center of Excellence (CoE) for an international automobile-manufacturing firm. Many companies are implementing the model where the CoE is responsible for Pattern definition, tools, practices, and providing guidance to application teams. Attendees were presented with the case study of how this issue was resolved by this an international automobile manufacturing firm.

**Working with the Dublin Core Metadata Standard**

*David White  
Lead Architect, Metadata  
BellSouth Corporation*

The Dublin Core is fast becoming the standard representation of metadata that for representing the semantic meaning of web-based documents. This presentation introduced the basics of the Dublin Core initiative and how organizations can implement this standard.

- Dublin Core Schema, Elements, Qualifiers, Expansion
- Dublin Core within HTML (Metatag)
- Dublin Core within XML (RDF)
- Dublin Core Syntax
- Recommended starting points for working with the standard

**Metadata and IT Service Management**

*Charlie Betz  
Technical Consultant  
Best Buy Corporation*

What is configuration management? What is the difference between incident and problem management? What is the purpose of a capacity planning capability? These IT process areas

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

have increasingly formalized definitions, based on the emerging general area of IT Service Management and its flagship standard, the UK-based Information Technology Infrastructure Library (ITIL). ITIL has been sweeping US Fortune 500 corporations over the past 2 years and, while it never mentions the word "metadata," calls for something termed a "Configuration Management Database" (CMDB). What is the relationship between this and the concept of a metadata repository? There are significant overlaps and opportunities for both synergy and disharmony.

This presentation discussed the ITIL concepts as they relate to metadata. In particular, the highly generalized data model of the ITIL CMDB concept was critiqued, and richer, more accurate approaches discussed.

## **Enterprise Information Architecture for Unstructured Data**

*Michael Reed  
Principal Consultant  
Enterprise Warehousing Solutions, Inc.*

This session discussed the integration of an unstructured data framework within an overall Enterprise Architecture. Efforts currently underway at the Federal Bureau of Investigation were used as an example of building a basic framework for the customer's understanding of unstructured data issues.

Unstructured data in the business, information, data, systems, and computer architecture were covered with examples of current successes and shortcomings. An overview of security requirements at the document level and their effect on table population was explained. The audience was surprised to learn that surveys show any government agency building an enterprise-wide taxonomy should expect over one million categories. Ontologies, glossaries, and data dictionaries were also briefly discussed.

## **Data Privacy and Security: What It Is, Why You Need It, and How To Do It**

*David Schlesinger  
Intel Data Policy Governance Manager  
Intel Corporation*

New data laws require an understanding of who is using what corporate data. These laws, Ordinances, Regulations, and Policies cross all departments and are a corporate issue. If each department tries to manage data regulations individually it will exacerbate system complexity, add data definition confusion, and inflate costs. Since data laws and policies are proliferating, we need an extensible way to manage them all. Thinking of "Regulated Data" as a class allows you to define sub-classes to meet present and future regulatory requirements. Data Administrators need to lead the way. Here's how:

- Focus on solutions to share and re-use enterprise data according to data regulations.
- Simplify redundant and costly account request processes into one standard and informed system.
- Drive standard data regulatory classification standards in metadata to reduce complexity, inform the business, and comply with data regulations.

**SPECIAL INTEREST GROUPS**

**Wednesday, May 5**

**7:15 am – 8:15 am**

**Outsourcing of Data Management: Will You Be Next?**

*Michael Scofield  
Assistant Professor  
Loma Linda University*

"Outsourcing" is one of the most ominous words in the business lexicon because it typically means that people are losing their jobs. And as the pace and reach of outsourcing increases in the data management area - for everything from database design, to real-time monitoring of systems, and now the latest trend of outsourced analytics and BI - then many of us are asking ourselves, could I be the next in line? Mike Scofield led this discussion of outsourcing trends and where they will go next.

**Post-Secondary Education in Data Management**

*Deborah Henderson  
Data Warehousing and Reporting  
INERGI LP*

*Anne Marie Smith  
Practice Director, Ed. Svcs. and Info. Mgmt.  
Staccato Consulting Group, LLC*

The DAMA Education/Curriculum team presented the work to-date on the DAMA-sponsored IRM/DRM curriculum. The session was well-attended, despite the very early hour. Members of the Education / Curriculum team - Deborah Henderson, Patricia Cupoli, Anne Marie Smith and Eva Smith (committee members Brett Champlin and Jeff Hoffer were not in attendance at the conference) presented the rationale behind the development of a curriculum for IRM/DRM education, its current focus on the North American educational system (2-3 year colleges, 4-year colleges, graduate schools), explained the structure of the proposed curriculum, and outlined some of the coming work for the committee (finalizing the current version of the curriculum's outline, presenting it for approval, marketing/promoting the curriculum). After the formal presentation, the attendees were invited to contribute questions, comments and suggestions. This led to a lively discussion and some attendees offered their services to the committee. The committee hopes to present the curriculum formally at the DAMA International Conference in 2005.

## CONFERENCE SESSIONS

Wednesday, May 5

8:30 am – 9:30 am

### **What do you do when the CIO says "forget the gurus - go save me some money!"**

*David Newman*

*Senior Director of Enterprise Data Management*

*Radio Shack*

David Newman is creating a comprehensive information asset management environment that will challenge some of the conventional wisdom around the role of today's data management organizations. He is responsible for defining and implementing a comprehensive information management program in a tight economic climate and an increasingly competitive industry. He is introducing a new value proposition for Enterprise Data Management, including:

#### **Compliance**

Teaming with Internal Audit to help the CIO and CFO, especially in areas such as SARBOX: How to assist Internal Audit functions with: Meta Data Management; ETL; Data Quality Management; and Data Model Management

#### **Helping the CIO on Operational Efficiency**

Identifying inefficiencies in current application portfolio, such as redundant point-point interfaces and how to create common views for a Single Version of the Truth.

#### **Architecting a better way - EDM support for EAI**

Migrating to lower cost platforms, establishing a common data layer and EDM's role in Enterprise Data Warehousing

### **UML for Database Design**

*Terry Quatrani*

*UML Evangelist*

*IBM*

The Unified Modeling Language is the standard language for visualizing, specifying, constructing and documenting all of the artifacts for a software system. Database modelers can now reap the same rewards from using the Unified Modeling Language as application developers have been doing for years. Application and database modelers can now speak one language - UML. Also, being able to link together the object and data models, and thereby improving the understanding of both, helps yield higher quality systems.

### **Success in a Storm: A Case Study on Wyeth Consumer Healthcare Data Architecture**

*April Reeve*

*Data Architect*

*Wyeth Consumer Healthcare*

While many corporations were trimming or eliminating their IS Data Architecture functions during these last two years, Wyeth Consumer Healthcare continued to invest and grow its Data Architecture team. Data Architecture had become a key contributor and trusted advisor to their

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

client organizations while other Data Architecture organizations have struggled to prove their value within Information Technology and their corporate organizations. This presentation discussed the factors and behaviors that have made Data Architecture a success at Wyeth Consumer Healthcare.

## **Meta Data Exploitation for Financial Gain**

*Christine Craven  
Customer Insight Development Programme Manage  
Abbey*

Increasing competition, strategic partners, mergers and acquisitions represent some major challenges for the organisations of today. However, effective development and exploitation of meta data can significantly facilitate information integration requirements associated with these challenges and at the same time provide substantial financial benefits and improved customer service.

This presentation provided practical examples of how Abbey has developed and exploited its meta data to improve its Cost : Income Ratio and eliminate some major customer service eyesores.

## **The Ins and Outs of Semantic Integration**

*JP Morgenthal  
Chief Services Architect  
Software AG*

Semantic integration is a powerful superset of data and systems integration. It allows us to integrate data and systems behind an abstract representation that provides additional context. JP discussed how semantics allows machines to understand how data and systems relate to one another and then provide abstraction, reusability and agility for future application development. He demonstrated the critical importance of metadata and explain how ontologies improve the capability for semantic integration by providing classifications that help to create order and improve the understanding of relationships between entities.

## **XML vs Relational: Comparisons and Contrasts**

*David Plotkin  
Data Quality Manager  
Well Fargo Bank Consumer Credit Group*

This presentation compared and contrasted XML vs. Relational data management and technologies. Many of the differences between them are quite clear, and their advantages and disadvantages obvious. However that doesn't necessarily make it easy to choose between the two, as trade-offs will always exist. David Plotkin identified the principle differences and similarities so that the audience could better understand each one and make better architectural decisions in future.

## **The Human Side of Data Modeling: Proven Techniques for Improving Communication With Subject Matter Experts**

*Alec Sharp  
Consultant  
Clariteq Systems Consulting Ltd.*

The human side referred to understanding and taking advantage of typical human likes, dislikes, etc. to make modeling more effective. We need to do this because:

- Time is more precious than ever, so we must help our clients quickly become comfortable and productive with modeling
- The human aspects of modeling can't easily be outsourced or offshored

Seven specific human side behaviors that modelers can use are -

- Accessibility – make it easy for non-modelers to get involved
- Directionality – models are clearer when drawn with a sense of direction
- Simplicity – get out of the detail into high-value conceptual modeling
- Consistency – do the same things the same way every time
- Visibility – don't hide patterns or symmetries in randomly drawn models
- Relevance – relate the model to familiar artifacts, business issues, etc.
- Plurality – use multiple techniques and appeal to multiple styles (visual, auditory, kinesthetic)

### **CONFERENCE SESSIONS**

**Wednesday, May 5**

**10:00 am – 11:00 am**

## **Challenges for Data Architecture in a Package Environment**

*Dawn Michels  
Senior Data Architect  
Guidant Corporation*

The data architecture discipline has relevance and value to add in an integrated application environment (IAE). Just because a corporation chooses to purchase "best of breed" packages to solve some application challenges, does not relieve the Data Architect of the responsibilities to ensure that the needs of the business are reflected in the purchased package.

The highlights of this presentation included the following:

- A DA can help provide value by defining business requirements and develop the data modeling artifacts of conceptual, logical and physical models to illustrate the requirements.
- Participate in the RFP process early – especially by providing criteria for the evaluation that is data focused.
- Do your utmost to secure a copy of the physical data construct from the vendors, and begin to map the data to the design of your data models. Document this mapping in an xls (format and make it available to all).
- Above all remain involved thought the lifecycle of the project. Put on your “scouting for the future” thinking caps, and point out opportunities for the future.

A couple additional ideas from the audience included:

- Be sure a member of the security team is involved
- Involve your DBA early in RFP so they are on board with what is happening
- Insist the vendor share their underlying proprietary data model – make it a “deal breaker” during negotiations.

## **Creating a Global Master Data Organization and Process: A Real World Case Study**

*Terry Haas  
Director, Customer Analytics  
Air Products and Chemicals, Inc.*

This session discussed how:

- Changing from a dispersed, decentralized data updating environment to a single, global master data organization can be accomplished.
- Success requires first defining a formal process and associated roles and then the organization.
- Key elements of a successful change plan include:
  - Data standards
  - Process and Quality Measures
  - Responsive Service Level Agreement
  - Clear Accountability and Control
  - Being part of a Business/Functional unit vs. IT
- Sustaining success requires pro-actively solving business problems – i.e. going beyond data maintenance.

## **Making Your Entities Behave: Entity Life Histories**

*David Hay  
President  
Essential Strategies, Inc.*

The characteristic of object classes in object models that distinguishes them from entity types in entity-type relationship diagrams is the specification of their behavior. It seems a reasonable extension to suggest, then, that when entity types representing things in the world are specified in requirements analysis, entity/relationship models could also be extended to represent the behavior of these entity types. The behavior of a real-world entity type, however, is often much more complex than the list of steps that could be captured in the lower third of a UML class box. So, while it is appropriate to want to know what functions are performed around the entity type, the description requires much more than a simple narrative. Enter the entity life history. This is a technique for representing the set of events that affect a particular entity type and the operations that are carried out in response to those events. It represents the complexity of different events and their effects.

## **End-to-End Enterprise Data Integration Toyota's best practices for integrated metadata, data migration and data quality management**

*John Gonzales, Elizabeth Kemper & Ken Karacsony  
Toyota Motor Sales*

In this session, Toyota Motor Sales described the best practices it successfully implemented via an integrated data integration suite – enabling them to incorporate metadata into data quality analyses, as well as data migration and transformation logic, and vice versa. Toyota explained its data integration architecture, and discussed the key benefits it has achieved since the implementation. These benefits include the ability to provide timely and critical information to business experts, as well as data integration and quality analysts. By capturing information about data modeling and architecture, source-to-target mappings, data transformation logic and data validations, Toyota is able to incorporate the results of its data quality analyses into its data migration logic and quickly address potential data anomalies.

## **Proactive Metadata-Driven Information Quality**

*David Loshin  
President  
Knowledge Integrity, Inc.*

"Active metadata" is a concept that is beginning to gain a developing stronghold in a number of organizations, as was evidence by some of the questions and comments that were made during my presentation on Proactive Metadata-Driven Information Quality.

As more data is used for multiple purposes across an enterprise, the perception of information quality depends on the information consumer and the corresponding application context. Fitness for use is directly related to compliance with the expectations of how data is used by the information client, and being able to measure compliance with those expectations can provide an objective assessment of the quality of the data. For any data set, many consumer data quality expectations can be refined into formal relational constraints to be incorporated as metadata. In turn, these constraints can be used as business logic for proactively identifying noncompliant data instances before they are integrated or committed. This workshop focuses on a process for successively refining naturally defined data quality requirements into formal metadata constraints and rules, as well as techniques for using these rules as a component of a proactive information quality assessment and monitoring program.

Questions that arose during the session involved the ability to capture state associated with the firing of information quality business rules, the ability to accurately state and transform information quality business rules, and how classification, object concepts, and taxonomies are increasingly relevant to information quality programs. Lastly, there were comments made about the important connection between information quality, data standards, and legislative/regulatory compliance.

## **A Checklist for Building a Taxonomy**

*Malcolm Chisholm*  
*President*  
*Askget.com, Inc.*

Taxonomies are more common than most IT professionals think. For instance, reference data tables often consist of classification schemes, even though they are not formally termed “taxonomies”. This presentation described a checklist of points that experience has shown to be useful when implementing a taxonomy. Three broad areas were covered:

- Design of the taxonomy
- Helping system operators use the taxonomy to accurately classify things
- Ensuring proper governance of the taxonomy, especially after production release

Examples of what can go wrong with taxonomies focused on misclassification of things by individuals using a taxonomy, over-expectation of what a taxonomy can deliver, and misuse of a taxonomy in analytical contexts for which it was not intended.

## **PANEL: Data Management Imperatives for Sarbanes Oxley Compliance**

*Bonnie O’Neil, Westridge Consulting*  
*Christine Crave, Abbey National*  
*Kent Graham, O’Melveny & Myers LLP*  
*David Schlesinger, Intel*  
*Simon Hankinson, PWC*  
*David Newman, RadioShack*

This presentation discussed the requirements of Sarbanes Oxley and how the data management organization can effectively help the organization meet its the compliance requirements. This is a new area to most DM professionals, but there are many opportunities to help the CIO, the Applications Services organization, and the Internal Audit Group with compliance.

RadioShack for example is focusing on 3 areas of support: 1) data quality (is the data on your annual statement accurate?); 2) data modeling and data transformation (what does the data on your annual statement mean?); 3) metadata management and governance/stewardship (where did the data on your annual statement come from and who’s responsible for it?).

Suggested questions to be addressed by the DM organization are:

- What services should data management organizations provide to their companies to help meet the compliance objectives of Sarbox?
- What role does data quality play in Sarbox?
- What role does data modeling and data transformation play?
- What role does metadata management and data stewardship play?
- What data governance policies need to be implemented for Sarbox?

**CONFERENCE SESSIONS  
Wednesday, May 5  
11:10 am – 12:10 am**

**Practical Experience with Enterprise Data Modeling**

*John Sykes  
Manager  
The Vanguard Group*

Since 1999, The Vanguard Group has actively pursued the deployment of enterprise databases. These databases now provide the back-bone for Vanguard's award-winning website, as well as new CRM applications being deployed company-wide. The ongoing Enterprise Data Modeling Effort has been critical to the successful roll-out of these applications. This presentation share those experiences, successes and pitfalls, including:

**Successes:**

- Adopting a maturity model for enterprise database evolution to establish a vision and temper expectations.
- Setting clear and measurable objectives for the data modeling team.
- Utilizing Reverse-Engineering to jump-start models.
- Adopting "structured agile" data modeling techniques.
- Embracing continuous process improvement.

**Pitfalls:**

- The desire for "top-down modeling".
- Dealing with conversion and interface teams.
- Analysis Model (Class Diagrams) vs Logical Data Model Dilemma.
- Insufficient Metadata for knowledge sharing.

**Enterprise Architecture: A Strategic Approach to Improving Data Quality and Restructuring Information Management**

*Jorge Yero  
Business Intelligence and Reporting Team Lead  
Aera Energy LLC*

This presentation described how Aera is using Enterprise Architecture to achieve data quality and fundamentally restructure the way it manages information. Enterprise Architecture Planning provided a rigorous assessment of the current state and clearly defined the target architectures. The subsequent implementation program is defined by an ambitious system development schedule and supported by a comprehensive Data Quality process. It is steadily delivering the desired state. The presentation discussed the essential components of Aera's approach and examined the critical success factors identified. Enterprise Architecture has proven to be a superior means to align business and IT, integrate people, process and technology, and reduce the costs of information management.

## **The Table Management Distribution System: Air Mobility Command's Reference Data Success Story**

*Charles Capples  
Senior Information Systems Engineer  
The MITRE Corporation*

*William Clark  
TMDS Team Lead  
Tri-Cor*

The Air Mobility Command (AMC) faced significant data integrity problems in the early 1990's. Individual communities within the Command developed command and control (C2) systems using their particular view of the data values. The communities did not coordinate with each other on standards so the systems had trouble sharing understandable data. The AMC managers decided that they needed to take several actions to improve the data sharing capabilities and the data integrity of their systems. A key component of this multi-faceted approach was the Table Management Distribution System (TMDS). TMDS provides consistent reference data across all 5 major C2 systems within AMC. In addition, TMDS distributes this reference data for synchronized updates. A process that used to take weeks now takes an hour for global distribution and application. The introduction of standardized reference data reduced the data exchange reject rate from 50% to 7%. TMDS now supports 125 systems with approximately 1400 tables of reference data.

## **Implementing Business Driven Information Management Practices: From Policy to Metadata**

*Wayne Harrison  
Senior IRM specialist  
BMO Financial Group*

Today's competitive business environment demands that you know your customer and understand your market. You need to ensure you have the right business information and can leverage it to make the right business decisions. Emerging regulatory requirements such as the BASEL II Accord, PIPEDA and SOX, are increasing the need for advanced information management (IM) to safeguard privacy and address information retention, auditability and legal accountability for information. BMO Financial Group (the Bank of Montreal Group of Companies) understands that information is a critical asset, and when managed appropriately, will support both these business and regulatory drivers.

This presentation described how BMO is building the foundation to manage information assets in these areas:

- Information Management Corporate Policy
- Information Management Framework and Governance
- Information Stewardship
- Business Driven Information Standards and Quality Programs
- Business level impact analysis through 'Pedigree' Metadata; tracing movement of data from source(s) to Target(s)
- Business Level integration of Legacy Systems & Vendor Packages; through Metadata

## **Efficiently Implement Data Lineage and Impact Analysis in 100 days**

*Gary Mersy*  
*Enterprise Data Manager*  
*Best Buy Co., Inc.*

Data Lineage and Impact Analysis are key deliverables of Data Professionals. Sarbanes-Oxley requires companies to “collect, process and disclose the information required in the company's [financial] reports.” Implementation of Data Lineage allows your auditors to validate the information on is accurate. Many IS shops have aggressive service level agreements to meet. Implementing Impact Analysis aids your development groups providing a roadmap for development and testing of systems. This capability can find hidden dependencies before that key job fails affecting your whole batch window.

Many companies don't implement a metadata repository because there is no perceived financial ROI. Best Buy has implemented a very successful metadata repository that provides Data Lineage and Impact Analysis. This session covered the common tricks, tips and traps to an successful and efficient Metadata Repository.

## **Managing Metadata in Oracle**

*Daniel Liu*  
*Senior Technical Consultant*  
*First American Real Estate Solutions*

There are several different ways of managing metadata in an Oracle database environment. Prior to Oracle9i, there are two popular methods to extract metadata from data dictionary. The first method involves querying the data dictionary using SQL statements. The second method involves running Export with ROWS=N, then running Import with SHOW=Y. Oracle9i introduces a new Metadata API (DBMS\_METADATA package). It allows database users to easily retrieve complete database object definitions (metadata) from the data dictionary. Attendees also learned the new way of managing metadata in Oracle9i using metadata API package tool.

## **Defining the Financial Value of Information**

*Robert Randall*  
*Senior Consultant*  
*The Intoinfo Consulting Group*

Times are tough for companies and governments. Information managers need to determine their organization's information value proposition should they want to convince their senior management to invest in expensive yet much needed initiatives. Robert Randall offered five critical success factors for communicating the value proposition to senior management and to the organization as a whole:

- Prioritize information system investments in a transparent and systematic fashion that demonstrates the greatest value to the organization;
- Understand the information risks of retaining too much information and having too little usable information - both of which are the result of poor information management;

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

- When planning for new initiatives, ensure information needs are aligned to business needs;
- Reconcile competing business and corporate information requirements; and
- Ensure that senior management understands your information value proposition by incorporating various costing methods including asset valuation, cost benefit, and financial indicators.

**CONFERENCE SESSIONS**

**Wednesday, May 5**

**1:15 pm – 2:15 pm**

**Information Stewardship: From Vision to Implementation and Beyond**

*Kimberly Martin*

*Sr. Data Administrator*

*Anthem Blue Cross Blue Shield*

This session provided attendees a complete guide for creating and organizing an Information Stewardship (IS) program. The speaker included details about how to manage the program, training, communications, and an IS web site demo. The following topics were covered and proved to be beneficial for business and Information Technology professionals planning to initiate an IS program:

- \* Implementation Strategy (Building a Business Case)
  - Conducting a Pilot
  - Securing Resources
  - Major Challenges
- \* Managing IS
  - Stewardship
  - Meta Data Management (Business & Technical)
  - Data Quality
  - Model Management
- \* Governance: Standards and Guidelines
  - Models
  - Meta Data
- \* Training, Awareness and Communication
- \* Building an IS web site

**A Contrarian View of the ODS Value in a CIF**

*Michael Bryan*

*Director*

*Nextel Communications*

Conventional wisdom on the role of operational data stores (ODS) in developing and managing a corporate information factory (CIF) has been a top down approach. The ODS has historically been an interim, staging area where CIF source data is staged and then transformed into a data

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

warehouse or data mart. Nextel's experience, however, suggests that the CIF's value and future may lie in its ODS rather than its integrated data warehouse. The speaker advocates a bottom up approach, and challenges the conventional implementations. The Nextel experience is summarized in the following eight misconceptions:

- Operational Data Responsibility - The ODS can serve as a data replication point among operational systems.
- Linkage - The ODS allows functional analysts to translate strategic decisions into operational actions, measuring and monitoring impact.
- Archaeology - The ODS allows lost or incomplete business rules to be re-discovered from the behavior of source systems.
- Requirements - By sourcing more data than required, the ODS can isolate the CIF from requirements changes.
- Business Case - The ODS can justify the CIF with less discretionary and more persuasive business cases.
- Risk- The ODS can convert the CIF's data quality challenge from risk to service.
- Impact - The ODS can add smaller tactical impacts from operational reporting to the benefits of the CIF.
- Sponsorship - The ODS allows a CIF's sponsorship to extend beyond an analytical champion.

## **Managing Your Project Manager: A Survival Guide for Data Professionals**

*Karen Lopez  
Principal Consultant  
InfoAdvisors, Inc.*

Karen Lopez led an interactive discussion about being effective as a data management professional on a systems development project. Points included:

- Earning the trust of a project manager
- Knowing both the system and the business goals of a project
- Responding to development needs while balancing enterprise needs
- How to deal with compromises
- How to leverage data management artifacts

## **How to Implement a Quick, Low-Cost Metadata Management Program:**

*Janice Tyson Wolf  
Information Technology Specialist  
Dept. of HUD*

*Diana Young  
FAA Data Registrar  
Federal Aviation Administration*

The President's Management Agenda (PMA) has directed all federal government agencies to improve the quality and reliability of government information and services. In this mandate, the President instructed agencies to share their knowledge and resources with other agencies. The challenge? How do federal agencies work within the spirit of the PMA and make the President's

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

goal a reality? What's more, federal budgets are dwindling and of course, we are expected to do more with less.

As information/data managers, we know the foundation for improved information/data quality is effective metadata management. In recent months, data managers from over twenty federal agencies have come together to address the PMA challenge.

## **Instituting the Meta Data LifeCycle**

*Barbara Nichols  
Principal  
Metaview360, Inc.*

Many different kinds of tools are used to create and maintain a data warehouse, such as: data modeling tools, data extraction, transformation and loading tools, data quality assessors, databases, gateways, business intelligence and reporting tools, web sites, and portals. Meta data occurs throughout the entire process, and with the use of each kind of tool. An effective meta data management lifecycle captures the meta data from all points in the process, organizes it for analysis, and makes it available to all the processes and people that need it. This presentation described a life cycle for meta data management, and the corresponding kinds of tool implementation architectures that support it, including:

- The Life Cycle Components - It's not an SDLC
- How Tools Enter the Mix - An Effective Tool Approach
- Dealing with Implementation Obstacles: Technical, Organizational, and Cultural

## **Data Management: The Next Big Thing**

*Tony Shaw, Wilshire Conferences  
Todd Stephens, BellSouth  
John Friedrich, My Krisalis  
Charles Betz, Best Buy  
Craig Mullins, BMC Software  
Neil Raden, Hired Brains, Inc.*

In this lively, fast-paced session, each panelist nominated his own version of Data Management's "Next Big Thing."

### **Charlie Betz**

Metadata management will converge with systems management and IT Service Management. It will not do so from a position of strength.

### **Neil Raden**

Traditional data warehousing is long past its useful life. The centralized design of data models for a whole enterprise never worked anyway. Expect to see data warehouse veterans moving quickly toward federation, grids, EII and non-traditional, non-relational backbones. Vendors of taxonomies for defining domains and semantic modeling will have their day in the sun. Moore's Law will finally be felt in data warehousing as customers abandon expensive, proprietary O/S's and hardware in favor of cheap, scalable resources and open source software. Not only will

data warehousing hardware and software become a commodity, the data warehouse itself will too. This benefits everyone as the focus finally finally turns to solving end-to-end problems instead of data, databases, data modelers and over-the-hill guru's with tired stories.

### **Todd Stephens**

We need modifications of the basic accounting principles of defining a capital asset. Accounting must alter its basic principles in order to account for data, information, and knowledge.

### **John Friedrich**

Meta data management will become more like data management. We are not going to have a single repository (just like we are not going to have just one database). We are not going to "capture for use" all the meta-data in one integrated repository tool, and certainly not going to move all of that meta-data from source to source, or data tool repository to data tool repository. And, while important for the Meta-Data integration, etc., solution, a Meta-Meta-Data Repository or Meta-Repository is not going to be "the answer", only another tool.

### **Craig Mullins**

**WICCA** -- Things will be getting more **wireless**, more **intelligent**, more **connected**, more **complex** and more **automated**:

**Wireless** – pervasive computing and pervasive management

**Intelligent** – software with built-in ability to adapt

**Connected** – bridges being cobbled together for the islands of information, applications talking to one another, business and technology connecting to deliver true value to the organization

**Complex** – more bells and whistles added to our DBMS and management tools making them harder to understand without...

**Automation** – tools that use built-in intelligence to move toward self-management

Additional NBT ideas were taken from the audience, including:

- RFID – Radio Frequency Identification – poised to generate a deluge of new data, and give rise to enormous privacy concerns
- Privacy and security
- Outsourcing will continue, especially on automatable tasks

## **Optimizing the Value of Metadata for Business People**

*Hans Yeager*

*Partner*

*SH Group*

If you ask a businessperson if metadata is important to them, he or she will more than likely say "no". If, however, you ask them whether it is important to know how an element in a report was calculated, or to know how current the information is, or to know from where the information was obtained, you'll likely get affirmative answers on all counts. This presentation helped the audience understand what is involved in delivering a practical business intelligence solution and covered the following key areas.

- Metadata that is important to business people
- Approaches to managing metadata
- Idea places to manage business metadata
- Business intelligence software metadata integration

**KEYNOTE PRESENTATION  
Wednesday, May 5  
2:45 pm – 3:45 pm**

**Universal Principles of Integration**

*Len Silverston  
President  
Universal Data Models, LLC*

How can we move towards integration versus disintegration? Why are some organizations able to integrate their data and systems better than others? Len Silverston, best-selling author, will share experiences, universal principles, and recommendations regarding effective ways to integrate data and systems. This keynote address covered issues such as:

- Creating a fertile environment for data integration efforts. The principle of “Integration requires trust” was discussed as well as how to create a space where integration can happen.
- Gaining support for data integration efforts by offering business value. The principle “Keep sight of the goal” sheds light on how to provide ongoing value.
- Dealing with conflict, understanding differing perspectives, and managing ego. The principle of “First understand oneself, then understand others, then be understood by others” provides a context for resolving data management disputes.
- Learning from the past and re-using models that have worked. The principle of “Learning from experience is powerful” was discussed as well as how reusable models, such as Universal Data Models, have helped organizations move towards integration.

The goal of this presentation was to share a picture of what integration looks like, what disintegration looks like, and principles that you can put into practice to move towards integration.

**SPECIAL INTEREST GROUPS**

**Wednesday, May 5**

**5:30 pm – 6:30 pm**

**Universal Process Models - Re-usable, holistic templates for the "other" side of the coin.**

*Len Silverston*

*President*

*Universal Data Models, LLC*

Just as "universal data models" offer re-usable components for quickly developing data models and data warehouse designs, there are many common processes that occur throughout various enterprises. For example, most enterprises have processes regarding marketing, sales, order processing, shipping, invoicing, collections, and accounting. While processes will vary from one organization to another organization, there are many processes that are very similar and process models can be re-used (and/or customized) in order to provide additional perspectives, quality checks, as well as providing templates for developing process models more quickly. This session discussed practical, re-usable, process models to facilitate quicker development of process models for common functions. Practitioners will often develop an enterprise process model in addition to an enterprise data model. Why re-invent the wheel regarding common processes such as contacting sales leads or checking credit history?

**Net Centric Data Strategy SIG**

*Michael Gorman*

*President*

*Whitemarsh Information Systems Corporation*

There was about 10 persons attending this special interest group. The goal of the meeting was to present, discuss, and receive input regarding DoD's 7 Net Centric Data Goals. The US Department of Defense "Net Centric" Data Strategy is aimed at making data assets more visible, accessible and understandable to all the various stakeholders. The discussion was lively and very constructive. A summary of the comments for each goal will be constructed and will be submitted both to the U.S. Army and to a key individual within the Office of the Secretary of the DoD (Tony Simon). A further discussion was held regarding the need to have a more formal user group on "data issues" within the DoD.

## **CONFERENCE SESSIONS**

**Thursday, May 6  
8:30 am – 9:30 am**

### **Establishing Best Practices within American Century's Data Services Department**

*Kevin Phelps  
Data Analyst - Specialist  
American Century Investments, Inc.*

American Century Investments Data Services Department has been establishing its Best Practices through:

- The Introduction and Utilization of Work Flow between DA, DBA, and application areas; Data Modeling Standards, & Data Administration Metrics
- Creating and applying an I.T. Four Year Plan & subsequently the Data Administration Operating Model
- By defining and launching our Metadata Direction & XML Direction, and
- By participating in the Architecture Working Group, a forum for sharing best practices.

### **Driving Strategic Direction and Results for Enterprise Data Architecture at Intel**

*Gregg Wyant  
Chief Data Architect  
Intel Corporation*

The Chief Data Architect at Intel Corporation discussed the challenges of introducing Data Architecture as a major management initiative in a large and distributed manufacturing company. The major challenge is to illustrate to senior management the strategic advantages of establishing an Enterprise Data Architecture, simplifying tools and methods, managing information with the same care as the company manages other valuable assets. He also showed his CIO level why central administration of data can actually free businesses users to develop more creative and useful local applications, and add greater data security. Mention was made of the power of the Zachman framework to direct individual efforts, and Data Quality, to locate systems of origin and records of reference that could serve the needs of the company with consistently defined business information. Key points were:

- Establish Enterprise Data Architecture org and Chief Data Architect role
- Base Enterprise Architecture on Zachman Framework
- Mandate Common Tools and Governance Process across organization
- All data (and its associated metadata) aligned to the Enterprise Model
- Data security centrally administered; data access centrally controlled

## **Putting "Place" in the Database - Spatial Information Storage and Use**

*Paul Baynham*  
*Senior Systems Analyst*  
*PlanGraphics, Inc.*

Location information is no longer consigned to the specialized realm of Geographic Information Systems software packages and gurus. More and more mainline commercial RDBMS support datatypes and storage solutions for location information. We as data managers will be exposed increasingly to these new data sets that describe location in new ways, and we will need the tools and systems in place to manage these data. In general, most location information has been geographic, thus the acronym GIS. But location applies to many things, and databases will be storing and using information about many things besides geographic locations. The mapping of the human genome uses an Oracle datatype to store the gene locations. Virtual buildings can have their layouts stored in spatial coordinates independent of any geographic system. New York City's non-emergency 311 telephone service uses location information to respond to callers' queries about the nearest library, park or police station. Once these types of information are stored in databases, then the job of spatial data management falls to the IT department and DBA. This is an exciting new element in the field of data management, and one to which not many of us have yet been exposed, but we will be more and more. It is time to get ourselves up to date on location data management and the metadata without which its use will be so limited.

## **What Your Mother Never Told You about Meta Data (Including, No One Makes Their Own Bootstraps!!)**

*Richard Weimar, Jr.*  
*Senior Consultant*  
*State of Michigan*

We have just successfully completed the first phase of the meta-Data Repository for the State of Michigan's welfare organization, The Family Independence Agency (FIA). We have progressed from being "green as gourd guts" to relatively sophisticated journeymen about what meta-data is, its true value and how to manage it to benefit the business. We were surprised by the significant potential value and power in good meta-data of broad scope which we had not considered initially. We are beginning to take advantage of these attributes. The key is - No one does this successfully by themselves. Other people have good ideas too! Accepting this up front and using it as the driving principle will significantly contribute to a successful project.

## **Refactoring Metadata: Finding Architectural Compatibility through Structural Comparisons**

*Baden Hughes*  
*Senior Research Associate*  
*University of Melbourne*

A plethora of metadata standards are emerging, often founded in different descriptive traditions, and documenting an increasing variety of objects. As a consequence, there is an increasing need to be able to compare, merge or translate metadata between different standards. Prior to

the conversion between different metadata standards, an interoperability analysis must first be undertaken to ensure that the correspondences between XML architectural components (namespaces, schemas, DTDs) and XML documents (elements, nodes) are understood and any incompatibilities resolved. Methods for such analysis are often complex, and require reference to formal standards in order to clarify structural relations. With the growth of metadata catalogues, there is increasing importance of automated methods for such large scale analysis and conversion tasks.

## **The Return of the Data Warehouse: This Time It's for Real**

*Neil Raden  
President  
Hired Brains, Inc.*

- Data Warehousing was a flawed concept from the beginning and the relentless progress of Moore's Law is placing enormous pressure on it to change or get out of the way. Most companies will need to scrap or wrap their DW's in the next 3-5 years
- The abundance of computing resources and bandwidth argues against excessively careful design and data modeling for performance, including data warehouses themselves
- Professionals should begin to refocus on more productive activities, such as broadening the scope of Business Intelligence or developing amazing software to manage the physical aspects of BI such as ETL, Materialization, federation, tuning, etc
- Vendors like Netezza and Ab Initio are disruptive (in the sense of Clayton Christiansen) and will cause established vendors to morph or fail
- Metadata in data warehousing is mostly ineffective, used for CYA documentation or, once in a while, impact analysis. It needs to be put in motion by new models of abstraction that are powerful middleware that hide all physical aspects of the system
- Role-based security is still a manual process, because the investments in metadata have failed to yield a rich enough metamodel that roles can be directly related to domains

## **PANEL: How to Get Business and Data Professionals Interacting**

*Larry Dziedzic  
John Zachman  
Michael Brackett  
Jorge Yero  
Larissa Moss*

There is a major chasm at Row Three in the Zachman Framework: a major disconnect between the information that business needs to operate successfully and the databases managed by IT. The growing disparity of the data resource is severely hampering the growth and success of most public and private sector organizations. Basic principles of modeling data to meet business needs, proper normalization of the data, and the real power that can be supplied by database management systems seem to have been forgotten. Business and data management seem to be on separate and diverging paths.

This panel considered questions such as:

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

- What is needed to close the Row Three chasm and build a high-quality data resource that meets the ever-changing information needs of the business?
- How can a high-quality data resource be built that will supply high-quality information to an intelligent learning organization?
- What needs to be done to return to the basic principles that have already been established?

**CONFERENCE SESSIONS**

**Thursday, May 6  
9:50 am – 10:50 am**

**The Agile Data Method: A Roadpath for Modernizing Data-Oriented Activities**

*Scott Ambler  
President and Senior Consultant  
Ronin International, Inc.*

- Modern software development processes are evolutionary (iterative and incremental) in nature, and many are agile
- Agile software development is not a fad, it works and is here to stay
- Data professionals must adopt evolutionary development techniques to remain relevant, and better yet they should strive to be agile
- Evolutionary data techniques such as agile model driven development (AMDD), test driven development (TDD), and database refactoring are described at [www.agiledata.org](http://www.agiledata.org)
- Data professionals and developers are relatively ignorant of the techniques of the other, but they must learn these techniques if they're to work together effectively
- You need to choose to work effectively with developers, otherwise they can and will ignore you

**How to Make Data Quality Indispensable**

*Paul Nettle  
Data Cleansing Manager  
UK Ministry of Defence*

*Sam Phoenix  
Data Cleansing Manager  
UK Ministry of Defence*

The Cleansing Project (TCP) was set up with a finite life expectancy to clean Logistics data for the Defence Logistics Organisation – and then disband. But by working closely with the supply chain business units, TCP were able to change business processes and system controls to reduce data corruption, and were able to demonstrate, through metrics and real-world examples, that they had played a central role in reducing waste and inefficiency within the Supply Chain. Over two years the team became the central focus for information quality within the UK Ministry of Defence, and by challenging the organisational culture of the Defence Logistics Organisation and by spreading awareness of Data Quality the team secured the

funding and senior management endorsement to change the team from a finite-lived project into a permanent Centre of Excellence in Data Quality.

## **Business Rule Engines: Build or Buy?**

*Roland Berg  
Principal Consultant  
ThinkSpark*

Several vendors have entered the business rule product space, all promising to deliver the management and implementation of business rules for our applications. This presentation looked at the approaches of some of the tool vendors and explore some of the issues to consider when making the build or buy decision. Strengths and weaknesses of the various approaches to the business rule problem were addressed.

## **Meta-data for Hub-Based Enterprise Application Integration**

*Jim Fulton  
Associate Technical Fellow  
The Boeing Company*

Enterprise Application Integration for large enterprises rapidly becomes too complex for point-to-point interfaces. The alternative is a hub-and-spoke architecture with hub mediating the flow of data among applications. With this strategy, a number of integrations technologies gravitate toward the hub: integration brokers, message brokers, data warehouses, web services, security managers. Managing the assembly, deployment and evolution of hub technology, as well as that of the constantly changing set of spokes, requires not just significant quantities of meta-data, but a meta-data architecture adequate to the needs of many people, with many perspectives, many goals, many tools, many asynchronous tasks.

## **Service-Oriented Architectures: Applying Data Modeling Disciplines and Techniques to the New World of Web Services, XML and Semantics**

*Fran Clark  
President  
Arpeggio Technology, LLC*

XML is becoming firmly entrenched as the standard for “data in motion” among e-business applications and in service-oriented architectures. As a result, important data is now moving across physical networks and across semantic boundaries instead of inside traditional data repositories. Creates new challenges, including intensive new requirements for visibility and understanding of data at both the technical and business levels. By leveraging XML meta data and other technologies, it is possible to create dynamic, extensible semantic models that vastly improve data effectiveness and manageability across XML service networks. This session discussed:

- How to apply both new and established data-modeling disciplines to the “naturally disordered” world of XML service-oriented architectures

**DAMA Symposium + Wilshire Meta-Data Conference  
Los Angeles, May 2-6, 2004**

- The role of semantics in modeling data in motion
- How new technologies, e.g., RDF, enable creation of semantic models
- The role of relational databases in the management of data in motion

**Evaluating DBMS Platforms for a Next Generation Data Warehouse**

*Michael K. Miller*  
*Manager, Data Management Group*  
*Cardinal Health*

In the last two years, Cardinal Health embarked on an extremely aggressive implementation of a new DBMS platform for its Enterprise Data Warehouse. A critical first step was selecting the DBMS platform that would serve as the engine for an advanced, active data warehouse implementation. At Cardinal, the data warehouse is a mission-critical component in the overall business and technology infrastructure. Consequently, it was imperative that the evaluation team submit the vendors to a rigorous yet objective process. This presentation explained the evaluation methodology and provide the audience with guidelines for establishing objective, bias-free criteria. Cardinal used a process that included performing benchmark studies with the vendors under consideration and this proved to be critical in crystallizing the decision for DBAs and management.

**Business Semantics and the OMG Business Rules SIG**

*Donald Chapin*  
*Business Vocabulary & Rules Consultant*  
*Business Semantics Ltd.*

In June 2003 the Object Management Group published an RFP requesting proposals for a metamodel which can be used to specify unambiguously the meaning of the vocabulary and rules of the business of any kind of organization. This RFP, "Business Semantics of Business Rules", requires that the metamodel be conceptualized and expressed the way people in organizations think and talk, independent from any reference to any information system. It is to be based on First Order Predicate Logic to enable largely automated transformations to information systems designs. Initial proposals were submitted for this OMG RFP in January 2004. This presentation discussed each submission, how it fits the requirements in the RFP, and the potential benefits to organizations from using methods and tools that support this upcoming OMG standard.

**SEMINARS**  
**Thursday, May 6**  
**2:00 pm – 5:30 pm**

**Being Right Isn't Enough: A Crash Course in Consulting Skills**

*Graeme Simsion*  
*Senior Fellow*  
*University of Melbourne*

The workshop covered the importance of communicating and working effectively with stakeholders (in addition to having the "right" answer) and provided some important principles and techniques. Key ideas included:

- identifying the full range of stakeholders (economic, user, technical)
- setting and managing expectations from the beginning
- clarifying the nature of the consultant-client relationship
- understanding and dealing with conflict
- the value of follow-up beyond the "end" of the project

The focus was on giving "internal" consultants some of the tools and techniques that external consultants use.

**Managing IT Professionals Seminar**

*George Flanagin*  
*Partner*  
*Bright Crayon*

**Managing workplace culture of IT:** IT professionals are more concerned than most workers about what their peers think about their skill level. Additionally, IT professionals quite often work at multiple sites, and intermingled with users, customers, and peers. A successful workplace culture has much to do with understanding the unique ways that IT professionals interact with each other, and helping them to understand the language of the users and customers.

**Listening:** Because a group of IT professionals is likely to have a different psychological profile (such as Meyers-Briggs Type Indicator), managers of IT professionals must learn to listen in a different way. In our look at these differences, we also discuss "emotional intelligence" and how it affects the management/IT interface in your corporation.

**Handling difficult situations:** IT professionals "rebel" more than other groups, frequently in response to issues of process and freedom.

**Managing the time of the IT professional:** The stereotype programmer comes in late, works through the night, and sleeps under his desk. This makes it difficult to track accountability, as well as to support the necessary communication with other groups. You need to create an orderly work environment that still has freedom.

**Identifying resources:** Most IT professionals are concerned with developing their careers, simply because the medium that they are working in changes quickly. The desired direction of career development may not always be in the direction the corporation perceives as "up."