

Conference Trip Report

2003 Enterprise Data Forum

Hilton Philadelphia/Cherry Hill • November 3-6, 2003

Conducted by Wilshire Conferences, Inc.

This report compiled and edited by Tony Shaw, Program Chair, Wilshire Conferences

Questions and issues may be addressed to - tony@wilshireconferences.com

The 2003 Enterprise Data Forum (EDF) was held before an audience of over 350 attendees and speakers. To receive more information about this conference, and related future events, go to <http://www.wilshireconferences.com>

This report contains a chronological summary of the key discussions and conclusions from most of the tutorials and conference sessions.

Reproduction Policy

This document is © Copyright 2003 Wilshire Conferences. It may be copied, linked to, quoted and/or redistributed without fee or royalty provided that all copies and excerpts provide attribution to Wilshire Conferences and the appropriate speaker(s). Any questions regarding reproduction or distribution may be addressed to info@wilshireconferences.com.

TUTORIALS

Implementing Universal Data Models to Integrate Data

Len Silverston

President

Universal Data Models, LLC

The Universal Data Model for Parties, Roles, and Relationships provides a solid foundation for data integration and allows data about people and organizations to be stored in one consistent place and a complete profile of all the roles and relationships for each party.

This model may be implemented in many different ways, for example, by allowing a single PARTY table, separate PEOPLE and ORGANIZATION tables, a single PARTY ROLE tables, separate tables for roles such as CUSTOMER, EMPLOYEE, PARTNER, a single PARTY RELATIONSHIP table, and/or separate relationship tables such as EMPLOYMENT and CUSTOMER CONTACT RELATIONSHIP.

In order to populate the tables, a "system of record" strategy and "pattern matching" strategy is needed. The integrated data store identifiers (such as party_id) need to cross reference the application keys and there are three main database structures that can accommodate this: placing a foreign key in the application table, building a cross reference table between the enterprise key and the application, or a combination of both.

The architecture for implementing these integrated structures may be virtual (for example common XML schemas) or physical (for example using an operational data store)

XML Schemas for the Data Architect

James Bean

President and CEO

Global Web Architecture Group

XML Schemas (W3C) are becoming the prevalent form of metadata for constraining an XML transaction or message. XML Schemas provide a robust set of rules and constraints that in some cases go well beyond the typical capabilities of a logical or physical data model. Yet, XML Schemas present tremendous similarities and synergies with the traditional roles of the data architect. Of significant importance are:

- Data Containers
- Structures
- Data Types
- Data Type Facets
- Data Value Thresholds
- Valid Values

Also of importance are the robust capabilities of XML Schemas to support reuse. This allows the data architect to promote standard metadata, definitions, and rules throughout the enterprise.

Establishing and Running an Effective Data Stewardship and Information Quality Improvement Program

Kurt Allebach

*Executive Consultant/Principle
Caladesi Professional Services*

Poor quality data has a tactical impact on organizational performance causing process failures and low customer satisfaction. Many companies are implementing Data Stewardship and Continuous Quality Improvement programs. Key to success in these efforts are:

1. A thorough and agreed understanding of the various data stewardship roles within an organization and the critical relationships between them.
2. Development of an Enterprise Information Framework and understanding of how it interacts with the Enterprise Process Model.
3. Development of Domain Information Standards that serve as benchmarks for the data quality analysis process.
4. An understanding of the various infrastructure elements, such as metadata repositories and analysis tools, within a Data Stewardship program and the function they play.

Mastering Reference Data

Malcolm Chisholm

*Senior Consultant
Askget.com Inc*

Reference data is also known as code tables, lookup tables, and domain values. It is found in all databases and is highly shared within, and even between, enterprises. This presentation examined the special management challenges that are unique to reference data and options for meeting these challenges. Not addressing these challenges can result in data quality problems and misfiring of business rules, which are often driven by reference data values. Specific attention was given to:

1. Managing the extensive metadata that is linked to reference data, and reusing it to provide assistance to both developers and business users.
2. Understanding the link between business rules and reference data, particularly in assessing which rules are impacted by which values. Unfortunately, there is little support in this area from business rules technology vendors.
3. The different kinds of change control that need to be implemented to manage different reference data life cycles.
4. The basis for building a justification for the introduction of centralized reference data management in an enterprise.
5. The role of centralized reference data management in an enterprise, and the risks it may encounter.
6. Basic design options in building a repository to house and manage reference data.

A Fundamental Framework for Evaluating Data Management Technology and Practice

Fabian Pascal

*Technology Analyst, Editor and 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 tutorial provide 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.

Process Modeling Concepts

Marcie Barkin Goodwin

*President & CEO
Axis Software Designs, Inc.*

Process Modelers have a good idea of what it takes to create a process model – but they don't (or shouldn't) work in a vacuum. Business users and managers play a critical role in contributing to and evaluating process models, and it's important that they understand what a process model is communicating. This introductory presentation focused on empowering both new modelers and the non-modelers on the team - giving them a better understanding of the modeling process and the meaning of model components, allowing them to participate with confidence. Marcie explained:

- What is Process Modeling and why do it anyway?
- The role of the business expert
- Diagram context and decomposition
- Gathering Knowledge – Interviews and group facilitated sessions
- Data in a process model
- Process modeling standards and procedures

Converting a Logical Data Model to a Physical Database

Thomas Haughey

*President
InfoModel, Inc.*

This presentation described a step-by-step process for transforming a detailed (data element level of detail), properly normalized Logical Data Model (LDM) into a Physical Data Model (PDM). Two major data modeling steps/phases must have preceeded this transform: Conceptual Data Modeling, in which the entities and relationships of concern are carefully modeled according to well-defined rules which are unvarying; and Logical Data Modeling, in which detailed data elements are formed and defined to meet the detailed information requirements of

the enterprise, and then properly normalized (using the unvarying "normal forms" as the rules) into the entity/relationship structure, resulting in a strictly and properly normalized Logical Data Model, ready for transformation into physical form for implementation.

In transforming an LDM into a PDM, the "rules" which must be followed are in fact defined by the business, and are the physical constraints (ie., speed of performance, space, cost, physical placement/distribution, security and integrity protection) which the business desires to be met. Unfortunately, typically, many of these constraints are mutually exclusive, ie., if the PDM satisfies one of them, then it cannot satisfy another. The very methodical, organized step-by-step process described in the tutorial showed how each of these considerations is analyzed and incorporated as the LDM is transformed into the PDM to arrive at a "best balanced" PDM which will optimally meet the best balance of the physical constraints imposed by the business.

XML Prototyping - Models and Structures

James Bean

President and CEO

Global Web Architecture Group

The broad scale acceptance and adoption of XML as a method of describing transactions and messages (e.g. EAI, Web, OLTP) presents a tremendous opportunity for the Data Architect. XML is a robust, self-describing metadata language. Given the importance of metadata and standards, the Data Architect can add tremendous value to the process. However, while there are numerous affinities and similarities between data modeling and XML Prototypes, there are also a number of important differences.

This presentation included a number of techniques for modeling transaction and message oriented XML in a manner that resembles traditional data modeling. In addition, a number of key XML prototyping concepts were covered:

XML Structure Models

- Vertical
- Horizontal
- Component
- Hybrid

XML Architectural Container Forms

- Rigid
- Abstract
- Hybrid

Enterprise Metadata Implementation: Learning from “Best Practices”

R. Todd Stephens

Director of the Metadata Services Group

BellSouth

Enterprise metadata and EAI may seem like a strange relationship but the reality is that metadata is one of the most critical elements of a solid application integration effort. Todd discussed and provided a three year learning curve for the attendees of this workshop. The session reviewed seven perspectives of an enterprise metadata effort.

- 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. Interfaces, components, schemas, DTD, web services, systems, documents, web pages, metrics as well as the components of the traditional database metadata effort needs to be looked at from a different view. The organization that plans on implementing an enterprise architectures needs to take a long look at the data architecture and ensure that it includes a heavy dose of metadata.

Implementing a Message-Based Data Integration Strategy

Dave McComb

President

Semantic Arts

Service Oriented Architectures mark a major shift from the way systems have been built for the last decade. One of the most profound changes in how distributed databases will be integrated. Most modern systems will employ a Service Oriented Architecture which will consist of:

- Message Queues
- Message Brokers
- XML Messages
- Asynchronous and synchronous invocations

The partitioning of systems into applications will be drastically changed to accommodate shared enterprise services. One of the key disciplines to prevent this from becoming chaos will be "Enterprise Message Modeling" which is similar to Enterprise Data Modeling, but differs in that it concerns itself primarily with modeling that data that will be shared via messages to keep the applications and services in synch.

End to End Security for SQL Server 2000 Data

Morris Lewis

President

Holistech Incorporated

Because of new government regulations concerning terrorism, hacking, and privacy, SQL Server database administrators are responsible for ensuring the safety and confidentiality of their data no matter where it is used. Common business practices relating to authenticating users' identities and authorizing access to data are no longer sufficient to meet government requirements for protecting data, nor are they able to withstand even the simplest hacking tools available to attackers. New environments like Microsoft's .NET Framework do offer new ways to protect data, but they also add new complexity to the production environment as a whole. Finally, application design is moving away from creating monolithic programs that implement all services users will

need to a breaking apart of services and functionality into individual components that will often reside on multiple different computers, thus creating an environment in which data may pass through several transformations as it flows between the client and SQL Server. This tutorial covered techniques for building more secure environments for your data.

Information Modeling in a Changing Environment

Graham Witt

Senior Consultant

Consulting Insights

Change is an unavoidable feature of every system acquisition project, with requirements not only being refined during the course of the project but frequently evolving into something quite different. Agile methods have evolved in response to this situation but, since the tools and techniques generally available to information modelers do not provide much in the way of support for rapid change, it is tempting to dispense with information modeling as if it were an unnecessary and time-wasting distraction.

This presentation described a variety of techniques that enable an information modeler to respond to a changing environment and add value to an agile or conventional project in such an environment. Topics include:

- Reasons for change
- Managing model changes
- Global changes
- Model reconfiguration
- Consequential changes
- Documenting model changes for reviewers
- Incorporating changing models in documentation.

Building a Business Case for Enterprise Metadata

R. Todd Stephens

Director of the Metadata Services Group

BellSouth

Business cases are essential tools for strategic analysis and decision making, and for tangibly defining the expected costs and returns associated with a metadata project. A business case is a tool that works for a business to look ahead, allocate resources, focus on key points, and prepare for problems and opportunities.

Tie your business case to the priorities of the business strategy within your organization, such as improving the product, increasing revenue and profit, better customer focus, cost reduction or improving internal efficiencies. Examples of metadata business case drivers include:

- Reduce the cost of ongoing maintenance within the Technology Community by providing immediate access to critical system and application information.
- Reduce the time to market for common services by deploying reusable Metadata enabled frameworks.
- Enable the offshore outsourcing efforts by enabling Metadata Technologies and Asset Cataloging.
- Immediately accelerate revenue growth in the telecommunications market via increased package bundling with enterprise metadata.

Objective Data Quality Assessment

David Loshin

President

Knowledge Integrity Incorporated

Because data quality issues are relevant only within the business context in which inspected data is used, data quality levels can only be measured with respect to business data consumer expectations. Relying on subjective measurements determined by software vendors only provides a subjective assessment from the point of view of an external party with little stake in the ultimate project success.

Objective data quality measurement relies on metrics relating directly to how information is being used and how missed expectations impact the business. Once expectations are isolated and understood, we can define assertions that capture those expectations that are used for measuring how information complies with those data quality rules. These rules, which seed our objective data quality metrics, are knowledge-based metadata related to the data sets, suitable for incorporation into the metadata repository. This tutorial discusses the process of exploring information, identifying data quality rules, and isolating noncompliance as a sequence of stages:

Marrying SQL, XML, Web Services and Grid Services

Ken North

Consultant, Author, Speaker

The race is on between IBM, Microsoft, and Oracle to provide persistent data management for XML applications, for Web services, embedded applications, Web stores, grid services, and other software. SQL DBMS products will increasingly be judged on how well they support traditional tasks (such as transaction processing) while evolving to provide new capabilities (such as integrated business analytics). The latest releases of data management software from the big three vendors unite SQL with multidimensional and document-centric (XML) data and grid computing. Problems will continue to arise with interoperability, data aggregation, and data and application integration. For this reason XML, messaging, and Web services are becoming increasingly important.

Although IBM, Microsoft, and Oracle cooperate on important World Wide Web Consortium (W3C) specifications, they take divergent paths when it comes to development and deployment platforms. IBM and Oracle are actively involved with Java, Unix, and Linux. Microsoft developed the .Net framework, in part, as an alternative to Java and Linux computing. SQL, XML, and simple object access protocol (SOAP) are vendor-neutral technologies, but the software infrastructure, client APIs, and developer tools from each vendor tend to have a distinct Java or .Net flavor. IBM and Oracle see J2EE as the environment for building Web services, while Microsoft sees Web services development through .Net glasses. IBM, Microsoft, and Oracle are heavily invested in XQuery; and IBM and Oracle are heavily invested in grid computing.

Valuing Information and Knowledge

John Ladley

President

KI Solutions

An organization's information portfolio contains great potential value. Many information management departments want to demonstrate this potential value of information and knowledge to upper management. There are multiple reasons for this:

- Proactive justification for enterprise information management (i.e. their own jobs)
- Reactive risk management to reduce corporate exposure to information related risk

However, intrinsic value, or potential value, has no meaning to CEOs. Nor however, would a single CEO tell you information is NOT an asset to their organization. The investment in moving and storing information is astronomical. It would seem then, that information as an ASSET means that it has to appear somewhere on a balance sheet on an on going basis, not only as a result of a merger, or calculation of goodwill (FASB 141 142).

The quality of information within an organization affects the value of that organization. It affects the stock price. Most work to date on valuing information has been related to these intangible classifications. However, John discussed approaches to valuing information that focus on the concept that Information has no value unless it is actually used. As long as the data sits in a database, the capital invested in placing it there has not return. Like the winter coat, it is not valuable until needed. You can create value by applying the data to business decisions, or improving processes, or even re-selling the information. In other words, if professionals make decisions and take actions they are accountable then for some percent of profit.

An organization with valuable information-based projects would be able to take these projects requiring information and knowledge, and assign an increase in profit to professional INVOLVED in the project(s). For example, a project to clean up customer data results in an increase in client retention of 2%, and a bottom line delta of US\$10,000,000. The organization has 12,000 professionals. 2000 of these are connected in some way with the business processes and project used to improve retention. Measuring the delta per professional results in:

$$\frac{10,000,000}{2,000}$$

or US\$5000 contribution to profit per professional. The enterprise could impute an increase in value (after tax contribution to equity) and book that as an intangible asset.

Net Income Delta \$10,000,000

Less Taxes (50% rate) \$ 5,000,000 The Information Project contributes to equity, and new intangible asset of \$5,000,000 results.

The ACCOUNTING efficacy of this exercise is certainly not industry standard. But it can be used to make an effective business case. And certainly, if the intangible aspect is unsuitable, a Net Present Value could be developed for the \$5,000,000, and this number used as an indicator of the new 'asset.'

CONFERENCE SESSIONS

When Best Practices Aren't Good Enough

Katherine Hammer

President CEO

Evolutionary Technologies

One of the largest contributors to the failure of e-commerce and the general skepticism about technology can be attributed to a failure to understand the complexity and importance of data integration management to every IT initiative. While there is a general understanding that data integration is important - to business intelligence, enterprise application integration, customer relationship management, and so on - there is little attempt to treat data integration as an key factor in an effective enterprise architecture. This presentation reviewed a number of major problems with the way software evaluation is conducted and suggested how companies can overcome this problem. Specific advice included:

- From this day forward, do not purchase additional enterprise software UNLESS:
 - Products can be re-used across projects/application environments
 - The combination of software and methodology supports auditability and efficient change management
- Rethink your IT organization along common functional requirements
- One day at a time – forever

Enterprise Modeling and Metadata: Steps (and Mis-steps) from a Real-World Project

Ray McGlew

Director, Data Administration

IMS Health

IMS Health began a process to create a logical data model of the data it collects, manages, and distributes to clients. This model was designed to assist the company in consolidating many stove-piped applications in several countries. We have made significant progress in creating this enterprise data model, and have added goals to bring some organization to the process. We have also started a comprehensive Global Metadata Repository implementation to complement the model. This presentation outlined the activities the team undertook to keep management excited enough to keep funding this infrastructure project.

Using AI for Data Audit and Quality: Barclays Bank Case Study

Adrian McKeon

Managing Director

Infoshare Limited

Most IT systems cannot measure the accuracy of outputs:

- Does the system work and where is the evidence?
- Are business decisions based on garbage?

Small data quality flaws at the start of a project magnify into inexplicable defects in end user outputs. Objectives cannot be translated into measurable performance indicators and nobody knows why.

Barclays used artificial intelligence to audit trail the history of each data record at sub field, field and record level from the source system to the warehouse. Information is stored in a virtual data layer separate from the IT infrastructure. Audit trails made the workings of the IT infrastructure transparent and end users are able to validate output, identify errors and track back to fix them.

Using SAP's Business Intelligence Solution for Enterprise Data Warehousing

Kevin McDonald

CEO

Compendit

Kevin discussed the overall architecture and features of the SAP BW product. Among the key points he made in terms of how the world's leading organizations are using BW to implement enterprise data warehouses:

- SAP BW supports various architectural options including ODS, DW, and InfoMart layers
- SAP BI doesn't 'solve all problems'
- Co-existence (Hybrid implementations) with data warehousing infrastructure is prudent
- Successfully BW implementations start with business strategy to assure alignment

KEYNOTE: Database and Software Trends

Ken North

Consultant, Author, Speaker

Panelists:

Bob Bickel, *Bickel Advisory Service*

Geoff Brown, *Oracle*

John Goodson, *DataDirect*

Dean Guida, *Infragistics*

Ken and his panel discussed a number of major technical and non-technical trends that are affecting database and software environments today. These trends include:

- Pervasive computing: billions of devices producing petabytes of data
- Emergence of web services and messaging models for integration and communication. But progress is hampered by an abundance of competing standards, as well as concerns about reliability and security.
- DBMS products are increasingly multi-purpose platforms (the Swiss Army Knife Model)
- Grid services will emerge as a significant architectural model for large scale data management
- The strong emergence of open source software (Linux is now in 10% of the Global 3500)

- Security concerns, including information warfare, industrial espionage, identity theft and international business linkages
- Privacy and Legal Issues, including in the context of emerging global outsourcing practices

A Little Appreciation - Is it too much to ask for?

Graeme Simson

Senior Fellow

University of Melbourne

Panelists:

Karen Lopez, *InfoAdvisors*

Len Silverston, *Universal Data Models*

Joe Maguire, *Consultant, Author*

Christine Mandracchia, *American Reinsurance*

1. Agreement that there is a real issue with data modeling (and modelers) being properly valued, and that this can translate into applications being developed on poor foundations.
2. A number of suggestions from panellists as to how to promote data modeling; direct access to business stakeholders was considered the critical factor.
3. Experiences from audience and panel suggest that good data modeling is critical in agile methods; models once established are difficult to change, and 'general practitioners' do not produce sufficiently resilient models.
4. Data modelers should be integral players in project teams (rather than "arms-length" consultants in a data management group).
5. Evidence presented (Mandracchia) that formal data modeling in development translated into reduced development costs.
6. "Make yourself useful" (Maguire, Lopez) was offered as a key tactic for staying employed: more formally, data modelers were encouraged to broaden their skills, and (particularly) to develop a knowledge of their organization's database platform and approach.
7. Data modelers have a key role to play in a packaged software environment - not only evaluating packages but advising on tailoring and integration.
8. Given the move towards architectures based on messaging rather than a common data model, data modelers need to work closely with other architects and be prepared to offer advice at the detailed (attribute definition and formatting) level.
9. Data modeling skills in analysis and user communication translate readily into other analysis fields.

Introduction to Unstructured Data Management

David Raab

*Partner
Raab Associates*

Barry Graubart

*Vice President, Business Development
ClearForest*

Pandu Nayak

*CTO
Stratify*

Jose Colon

*Vice President of US Technical Client Relations
Autonomy, Inc.*

Unstructured Data presents challenges to traditional data management. It comprises the majority of corporate information, but is often difficult or impossible to access. This session explored the types of unstructured data, the challenges of managing it, and the technologies available to help meet those challenges.

Problems:

- Content is growing exponentially
- Typically stored in silos and randomly across the enterprise, not easily accessible
- Difficult to act on disparate pieces of information

Solutions include:

- Extend the data warehouse to unstructured data through tagging
- Rich metadata creation enables advanced analytics

Best Practices:

- Unstructured data management today is much more than simply search or taxonomy management.
- Tagging applies structure to unstructured data
- Once tagged, data can be integrated with structured data, enabling comprehensive business intelligence

Service-Oriented Integration and Process

Ron Schmelzer

*Founder and Senior Analyst
ZapThink, LLC*

Integration is not about simply plugging two systems or organizations into each other. The vision of "plug and play" application and system integration is a pipe dream that may be appropriate for the distant future, but right now enterprises face the more immediate challenge of connecting arbitrary systems in a manner that is cost effective, manageable, efficient and secure. Ron Schmelzer, discussed Web services and the Service-Oriented Architecture (SOA), as they

represent an approach for integrating systems using an abstracted methodology called Service-Oriented Integration (SOI). Major trends and objectives for organizations are:

- “Thrift” is the New Normal
- Web Services have taken hold during an IT downturn
- Reduce the cost of integration
- Squeeze more value out of legacy apps
- Embrace heterogeneity
- Increase business agility

UML for Database Design

Terry Quatrani

UML Evangelist

IBM

Application and database modelers can now speak one language - UML. No longer are database analysts, modelers, and designers relegated to the tail end of the development lifecycle. Now database designers can participate from the inception of the project, helping shape those early decisions that often have a critical impact on the system’s data. Also, being able to link together the object and data models, and thereby improving the understanding of both, helps yield higher quality systems.

The UML is the standard language for visualizing, specifying, constructing and documenting the artifacts of a software –intensive system. It will address ALL aspects of the software development lifecycle. It continues to be extended and refined by soliciting industry feedback and including the best constructs from other methods and languages.

In building a visual model of a system, many different diagrams are needed to represent different views of the system. The UML provides a rich notation for visualizing our models. This includes the following key diagrams:

- Use-Case diagrams to illustrate user interactions with the system
- Class diagrams to illustrate logical structure
- Object diagrams to illustrate objects and links
- Statechart diagrams to illustrate behavior
- Component diagrams to illustrate physical structure of the software
- Deployment diagrams to show the mapping of software to hardware configurations
- Interaction diagrams (that is, collaboration and sequence diagrams) to illustrate behavior
- Activity diagrams to illustrate flows of events

To Federate or Consolidate: 10 Things to Consider Regarding Data Integration

Ho-Chun Ho

President

HoTech Corp

There are many mechanisms to integrate data: data layer integration, data access layer integration, application-specific solutions, application-integration frameworks, workflow or business process integration frameworks, digital libraries with portal-style integration, search-engine-oriented integration, data warehousing, and database federation. When architecting data

in support of any of these integration options data architects often face a tough decision-- we need to choose between consolidating data and federating data. This presentation will share the insight including technological issues, organizational impact, as well as cost factors. Among other considerations, this presentation addressed:

- Basic principles and criteria for exceptions
- Data quality and integrity
- Transaction characteristics
- Performance, currency, and availability
- Abusive reuse of API's and middleware
- Structured, semi-structured and unstructured contents
- Ownership and turf wars
- Actionable meta-data

XML, Your Data, and You

Evan Levy

Senior Partner

Baseline Consulting Group

The presentation "XML, Your Data, and You" presented the benefits of using XML as a core technology of an IT environment. The industry analysts have found that 60% of development budgets are focused on data integration. The benefits of XML include code simplicity, increased integration functionality, and significant cost savings.

- XML isn't just a single technology, but a family of technologies that will greatly simplify the aspects of data access -- data manipulation, data display formatting, and data integration.
- The plain english structure of XML is of tremendous value. It's adoption in data extraction and load construction has greatly reduced data support issues associated with data marts and warehouses.
- New technologies (like EII) have embraced XML to enable a new means of integrating metadata management into database functionality.
- Through Xquery, product vendors have delivered a new generation of data access that provides a more efficient means for distributed database access.
- The importance of data management and administration is critical. XML can only enable application- and enterprise-based access if data management rigor is in place.

Data Profiling Technology: How to Recover Metadata and More

Jack Olson

CTO

Evoke Software

Data Profiling is the use of analytical techniques on data for the purpose of developing a thorough knowledge of its content, structure, and quality. The basic steps are:

1. Gather metadata
2. Identify and Prepare Data for Profiling
3. Value Analysis
4. Structure Analysis
5. Single Object Data Rule Analysis
6. Multiple Object Data Rule Analysis

Active Metadata

Adrienne Tannenbaum

President

Database Design Solutions

Panelists:

Naresh Govindaraj, *Informatica*

Greg Blumstein, *Data Advantage Group*

Andrew Manby, *Ascential Software*

Adrienne Tannenbaum started this workshop by illustrating today's naiveté when it comes to maintaining an "active outlook". She emphasized the need to keep the following parts of the metadata solution world current and up-to-date:

- The ROI - Most organizations start with an immediate metadata solution ROI (like reducing development costs via the provision of efficient impact analysis) and then forget about it once that initial objective is met. Successful metadata solutions evolve, and hence so does the ROI - the return should become greater over time
- The Organization's Development Approach - If the metadata solution is not an active part of application development, constant catch-up is bound to result!
- The Metadata Solution itself - metadata is active when it is always reflective of the instance of whatever is being described.

Vendors then presented overviews of how their specific products support "active metadata". An interactive Q&A then resulted, with attendees running out of time to continue their questions!

Name/Address Matching and Consolidation

David Raab

Partner

Raab Associates

Panelists:

Frank Dravis

Vice President Information Quality

Firstlogic, Inc.

Ramesh Menon

VP, North America

Search Software America

Michael Dunkerley

VP, Global Marketing Core Technology

Search Software America

Record matching is more than just name/address matching. It is important because identifying the relationships between customers, contacts, organizations, consumers, households etc within one system or database, or across multiple systems or databases, is a basic requirement of today's

systems....data consolidation, customer centricity, data cleanup, risk minimization, information reliability, sales & marketing efficiency.

Equally as important is identifying the relationship between the identity data in a transaction (or file) and the data stored in databases and files.....Customer identification, duplicate prevention, search reliability. Missing the data you have about a customer or prospect can mean bad business. Failing to successfully match against alert lists, fraud reports, and other important “screening” data can be disastrous to your business as well as have severe social impacts.

Searching and matching is a two step process.....first find the candidates that could be relevant to the search....second match the candidates to the search data to provide a confidence level or match result. Finding candidates in today’s data volumes require keys....some type of partitioning of the data that provides variable selectivity of the candidates in a search to (a) allow for the required real-time performance and processing turn-around and (b) limit the false match.

If the keys do not support finding all of the relevant candidates, there is nothing that the matching can do to uncover a match. If the keys, and the ways the keys and other data required for matching are stored in the database, do not support returning the results within the timeframe required, then the effectiveness of the results is diminished. The design of the keys is critical to the success of a search and matching system. The design of the database to hold the keys and other matching data is critical to the success of a search and matching system.

Maximizing search and matching reliability and quality requires:

- maintaining the integrity of the original data through the data capture processes
- retaining and using the original raw data regardless of the variation and error
- not believing the order, format or parsing
- use “smart” indexing that supports multiple keys, “many to many” indexes and the ability to find the relevant data regardless of the error and variation
- having available a selection of search strategies to fit the business purpose and transaction risk
- high quality matching processes that mimic the expert user’s ability to “overcome” the error, variation and presence of nulls
- scoring, ranking and presentation of results that simplify the user’s choice process or auto-matching
- consolidation processes that support high quality future discovery and do not discard knowledge
- tools that support multi-national data

A Comparison of UML Class Models and ERDs for Data Modeling

Paul Dorsey

President

Dulcian, Inc.

Many people question whether any part of the Unified Modeling Language (UML) can be used for data modeling. Some have suggested creating a new tool to explicitly support data modeling. However, with some extensions, the UML can be used very effectively to design databases. This session provided an overview of UML class diagram syntax as it pertains to data modeling and a discussion of how each drawing element can be implemented in a relational database.

One of the most challenging problems in mapping an object-oriented design into a relational database is how to implement generalizations. The traditional mapping of each class to a table generates logically correct but unusable systems. Redundant storage of inherited attributes along

the inheritance path is a strategy that allows modelers to use generalization without hesitation. The speaker also covered how logical Primary Key specification is still useful in class diagram data models and how the rules of normalization can be adapted to support object-oriented database design.

Exploring the Converging Worlds of Data Integration Tools

Faisal Shah

*Chief Technology Officer
Knightsbridge Solutions*

ETL or EAI toolset choices should be driven first by corporate meta-information culture, then with consideration given to existing toolset investments and architectural fit. Organizations need to address ETL and EAI holistically and at the same time understand that there are still significant differences between the tools and ways to approach integration projects -- each treat latency, unit of work granularity, meta data integration, third-party product integration, and other product dimensions differently. While EAI and ETL tools continue to grow closer together, there are still significant advantages to using each for its original purpose, for instance ETL is data-centric and meta-data driven and is excellent at bulk data handling while EAI is process-centric and event-driven.

- As a technologist, be careful not to fall in love with integration technologies
- These technologies tend to be weakest at the one activity that consumes most effort during integration projects
- Information technology is bifurcated into "system domains"
 - Data-centric
 - Primarily for analytical systems
 - Process-centric
 - Primarily for transactional operational systems
- The bifurcation often exists for good technical reasons but irritates business users by limiting which of their requirements can be met by IT

Differentiating strengths:

- ETL
 - Efficient bulk data movement
 - Highly refined meta information repositories and connectivity
- EAI
 - Low latency
 - Event-driven (very fine granularity events)
- Web services
 - Hierarchical service definition
 - Low latency
 - Event-driven
- EAI is becoming synonymous with Web services. This leaves two major integration tool space segments.
 - ETL for data-centric ideology
 - EAI and Web services for process-centric ideology
- Leading-edge vendors in both segments are targeting enterprise-wide integration capabilities with the primary distinction being system domain centrality
- What's missing in ETL?
- Forward thinking ETL vendors have realized their strength in meta data can be converted into a general meta-information strength

- But ETL vendors today are hindered by their legacy of delivering bulk data processing, and cannot deliver the fine-grained units of work and low latencies that EAI and Web services products can deliver

XML: A Component of Metadata and Enterprise Architecture

Ben Jenkins

Senior Architect

BellSouth

This session focuses on how the Extensible Markup Language (XML) applies to metadata, and offered an overview of one real-life implementation of an enterprise XML repository. As the universal language for data on the web, XML gives you the power to deliver unambiguous data meaning between business applications, search capabilities, and static/dynamic presentation.

The Metadata XML is a landmark in Data architecture evolution.

- It allows authors to say what they mean, rather than merely how to say it.
- XML carries the torch forward to a new generation of information services.

RDF Can be used in a variety of application areas:

- Search engines
- Intellectual property rights of web pages

RDF: the basic model consists of three (3) object types:

- Resources (the artifact being described)
- Properties (a specific aspect, characteristic, or attribute that describes the resource – these are highly extensible)
- Statements (resource + property + value of “property”)

Dublin Core: A standard set of “properties” to consider when identifying an RDF resource

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

How to Build a Great Relationship with your Data using Profiling and Quality Assessment Techniques

Brad Darrach

*Information Quality Analyst, IT Business Systems Support
Fallon Community Health Plan*

Unfortunately, most implementation projects have finite timetables with job implications.

The importance of data quality is increasing exponentially with the adoption of BI, AI, and other predictive modeling tools.

The amount of money you spend on a date doesn't necessary ensure a happy ending. You have to keep working at it. Whoever said that a job well done need never be repeated apparently wasn't married and never cleaned, painted, or managed data.

Tools for Success

- Identify Data Owners and Data Stewards
- Create Glossary of Terms – ensure data matches the definition and resolve ambiguous usage; include all sources and uses
- Know the data and look beyond the immediate source
- Involve customers in the design and implementation of edit checking and include in project business requirements
- Communication is key
- Metrics and feedback
- Audit and controls

Next steps

- Continue to “improve the process”
- Refine metadata – will be a moving target
- Improve statistical data surveillance
- Key Performance Indicators/Dashboards
- Data mining
- Establish true accountability for data owners

Integrating Data with Processes: Data Model-Driven Applications using CASE

Jose Borja

*Senior Analyst Programmer
Mayo Foundation*

This presentation used a live demo of a CASE-driven application to demonstrate:

1. How CASE applications:

- Are not hard coded
- Are portable
- Can be deployed to multiple platforms
- Can be shared and reused
- Are technology independent

2. Model-Driven Apps are an INVESTMENT because they:

- Capture data and business rules
- Capture the processes that deploy the business rules

- Capture the interfaces that implement the processes
- Are abstract and language independent
- Can be modified/redeployed across multiple languages and platforms with minimal labor expense

3. Summary Points:

- Information Engineering (IE) provides a “data centric” approach to software development in a CASE environment
- Models are an investment – Coded applications are an expense
- Data and process integration is a reality in a CASE environment with its own life cycle methodology
- Models are re-deployable to multiple languages and RDBMS's
- IT staff learns how to build models and diagrams...IT staff does not learn to write code in several languages

Enterprise Common Data Architecture - Roadmap to Integration

Daniel Paolini

Director, Data Management Services

State of NJ Office of Information Technology

The creation of an Enterprise Common Data Architecture (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.

GIGA Group defines Data Architecture as comprising “the vision, principles and standards that guide the creation, use and management of data and the deployment of data-related technology within an enterprise. The scope of data architecture includes the governance of all activities and processes involved in the definition, creation, formatting, storage, access and maintenance of data. While data warehouse and data mart architectures are important parts of a data architecture, these represent specialized subsets of the domain.”

NJ CDA Standards & Practices

- Data Management Framework
- Data Object Naming Conventions
- Enterprise Data Dictionary
- Reusable Metadata
- Iterative Data Warehouse Development
- Dependent Data Mart Development
- Standard Tool Sets and Platforms

NJ CDA Physical Components

- Spatial Data Services
- ETL Platform
- Business Intelligence Platform
- Database Platform
- Data Quality & Geo-coding
- EAI Platform
- Data Mining Platform
- Metadata Profiling Tool
- Metadata Management: Being Revisited
- Unstructured Content: Being Researched

General Advice:

- Which Tool is Less Important than Selecting One *
- Executive Sponsorship – Bully!
- Enterprise Funding – Be Creative
- Create Vision – The Plan Can Change!
- And Don't Forget Develop Your People
- Embrace Opportunities
- Eliminate Inbreeding
- Do Different with Less
- Herd Cats
- Be Firm but Flexible!!!

Enterprise Information Integration (EII): A Logical View of the Distributed Enterprise

Nitin Mangtani

Technical Program Manager

BEA Systems Inc

Business imperatives for real time information integration are increasing. Organizations looking to empower their employees with decision-driving information are faced with one common challenge: providing a unified view of critical business data. In today's extended enterprise, the time between when a piece of information is generated and the time when it is consumed is constantly shrinking. Delivering up-to-the-minute information to customers here not only increases the service level and therefore customer loyalty, it also decreases customer service costs

ETL/Data Warehousing

- Data warehouses are necessary to store large quantities of historical information, they are suitable for historical analysis typical of strategic planning exercises.
- Critical front-office applications such as customer service apps need information which is far more up-to-date than information that is available in the warehouse.
- Information needs in the operational arms of the business are far more dynamic than a warehouse can accommodate. If business needs information that has not already been planned to be staged in a warehouse, the cost and complexity to accommodate that change are usually prohibitive.

XQuery

- Preserve logical/physical data independence
- The semantics is described in terms of an abstract data model, independent of the physical data storage
- Consume/query data in its most popular interchange form – XML
- Declarative Programming - programs should describe the "what", not the "how"

Why a native query language?

- We need to deal with the specificities of XML (hierarchical, ordered, etc)
- XQuery is an emerging W3C Standard: <http://www.w3.org/TR/xquery>

Enterprise Data Maturity

Burt Parker

Principal

Paladin Integration Engineering

Peter Aiken

Founding Principal

Institute for Data Research/VCU

Enterprise-wide management of data is understanding the current and future data needs of an enterprise and making that data effective and efficient in supporting business activities.

Data management must answer to two sets of inputs: enterprise-wide data needs and functional user data needs.

- Enterprise-wide needs come to enterprise data program from the information management program in the form of directives, policies, and approved funding; enterprise data engineering coordination with enterprise levels of systems, technology, and process engineering; and coordination and liaison with enterprise data management activities of external organizations.
- Functional user needs for a particular business functional domain come to functional data engineering via coordination with their business function's business process engineering activities, feedback loops from data operations, coordination with functional engineering levels of other business functional domains, and coordination with functional levels of systems and technology, engineering.

Answering to both sets of inputs ensures that data management directives and goals, data models, data designs, and data assets fulfill both "top-down" as well as "bottom-up" requirements. The feedback loops ensure the continued viability of the data program over time.

Enterprise Data Management

- Purpose: Enhance the effectiveness of the enterprise by optimizing use of its data assets
- Implementing objectives
- Data Program Coordination: Provide appropriate data management process and technological infrastructure
- Enterprise Data Integration: Achieve enterprise data sharing of appropriate data
- Data Stewardship: Achieve business-entity subject area data integration
- Data Development: Achieve sharing of data within a business area
- Data Support Operations: Provide reliable access to data
- Data Asset Use: Leverage data in business activities

Semantics in Business Systems

Dave McComb

President

Semantic Arts

Semantics is the study of meaning. Most of what we do as data modelers and system developers is directly involved with the meaning of the information we are managing, and yet we spend little time with the study of the underlying discipline. This talk covered the eight levels of semantic precision that could be practiced in a given application:

- Label -- just naming an item with a label
- Tautological Definition -- a definition that adds little beyond what was already in the label
- "Good" definitions -- a definition that actually described the item
- Taxonomy -- the items place in a hierarchy
- Distinction -- a taxonomy with a way of distinguishing peers
- Ontology -- a network of related concepts, and the definition in that context
- Primes -- definitions built from a limited set of axiomatic primitives
- Formal semantics -- provable inference

Quantifying the Operational and Financial Risks in Enterprise Data Architectures

Chito Jovellanos

President & CEO

forward look, inc.

This conference session discussed the components of operational risk, with particular attention to the contributions from transactional systems and data. The evolution of the 'Basel II Accord' provided the framework for identifying specific external and internal risks, and methods for offsetting and hedging those risks. A summary of significant event loss data and their relative risk metrics were presented, spanning both high-frequency low-cost events (eg, transactional effectiveness) and low-frequency high-impact events (eg, natural disasters). Risk offsets include expensing, capital set-asides, outsourcing, insurance, and potentially, structured products (derivatives).

Data Modeling: A Developer's View

Roland Berg

Principal Consultant

ThinkSpark

Working with developers:

- Understand their perspective
- Understand how they are measured
- Understand how they measure themselves

Developer think:

- Task Oriented
- Wants to build cool stuff
- Get it done fast
- Micro view
- Generally only wants to know enough to get the job done

Developers main concerns?

- What do I need?
- How should it look?
- Where is it?
- How do I use it?

Teach Developers to Read Models

- Explain the purpose of the model

- Understanding
- Communication
- Explain notations

Explain the importance and meaning of the model

- to the business
- to the program/project
- to you the data architect
- to the developer

Important!!! Explain that each model type (conceptual, logical, physical) is one interpretation of the layer above and that while it must be a valid implementation of the model above it does not need to mirror it's structure.

Summary

- The more abstract your models the bigger the problem!
- Education and Communication are key
- Architects must take leadership role
- Burden is on the architect to ensure that the vision is properly communicated, understood and implemented
- Once they understand developers really like this stuff!

XML Vs. Relational - The Top Ten Differences

Jim Stewart

*Director of Consulting
ASIX, Inc.*

This presentation explained the fundamentals that differentiate XML and Relational technology. Although the differences between XML and relational seem obvious, the problem of selecting the right one it is not that simple when you take a closer look. Arguments for and against the use of the two approaches show confusion about the best uses for each, how they are different or similar, and apparently conflicting claims in several areas. The advent of XML DBMSs makes the issue even more confusing.

This presentation identified the top differences (and similarities) between the two technologies in a way that will help data professionals understand them, select the correct solution for an application and make sound architectural decisions on their use.

The Top Ten "differences" that reviewed were:

- Primary Purpose
- Data Paradigm
- Metadata Approach
- Ease of Understanding
- Relationships
- Object Identification and Navigation
- Query Ability
- Flexibility
- Data Integrity
- Performance

The Structural Integrity of Source Data

Joseph Novella

President

Anthem Consulting, LLC

Many data profiling and assessment efforts tend to focus on domain studies tasks and techniques. However, understanding the content and scope of individual columns and fields is only the first step in developing a complete data assessment.

Structural Integrity is the measure of an object's enduring ability to serve its designer's purpose, during changing and/or challenging conditions or stress, without failing or collapsing. Data's Structural Integrity means that:

- Rules maintained
- Logical consistency
- Controlled redundancy
- Minimal duplication
- Flexibility in satisfying business needs
- Over time
- With minor changes
- Minimal "kludges" or workarounds

Summary

- Documentation and metadata are not sufficient in quantifying the structural integrity of source data
- Functional dependencies reveal the inherent data structure, rules, and relationships quickly and reliably
- Review and prioritization of results is key to the process
- Continuous measurement and analysis of structure necessary

Positioning Metadata as a Business Asset (and get funding for your project)

Daniel Riehle

Principal

GetReals Inc.

Metadata is seen as a technical necessity, not as a business asset. This is a strategic error on the part of IT organizations. The reality is that funding for Metadata will never be significant if you can not morph metadata into a enterprise-wide business asset. This presentation showed how to use an extensible repository, COM objects, XML and ASP to generate self-guidance of business people through the mountains of technical metadata inherent in your IT organization. As an example of the technique, the presentation relies on the need to reflect "Lineage", the journey that business-valued information takes from first source (Legacy systems) through final targets (BI Tools).

To properly position metadata as a business asset (and hence secure funding for the IT effort), you must make the requisite detailed metadata transparent to the business user yet keep these users' need foremost in the overall architecture of the enterprise metadata. It is a sneaky detail that by focusing on the business value of metadata, you accumulate the technical metadata required to streamline the activities of IT.

Key take-away points:

- Extending the repository to reflect business needs
- Data acquisition and parsing of metadata sources (lineage and transform acquisition)
- Summarization of lineage
- Applicability to other enterprise metadata tasks
- Business and technical funding points (budgeting).

Role Transformation: Data Analyst and Architect

Jane Carbone

Partner

infomajic, llc

The role of the data architect is to provide the plan for enterprise data. This position works with the business (e.g., data stewards) to formulate data policies and plans that support enterprise goals, reduce costs and leverage the use of existing assets. This position works across the IT organization to ensure effective implementation, architecture compliance and conflict resolution. This position is a member—Architecture Governance Council. Reports to Chief Architect. The scope of this position includes all enterprise data (DW, databases, flat files, externally acquired data, etc.) with emphasis on mission-critical common data.

Responsibilities include development of:

- Information Policy that includes high-level principles around the value and management of the enterprise information assets
- Information Strategy that includes, for example, criteria for strategic (internally managed) Vs. tactical data (can be outsourced), description of desired use of data for OSS vs. MIS functions, articulation of data integration strategy, direction for data consolidation, description of target DBMS characteristics—describing “best fit” data store characteristics
- Data focus in architecture model, including identification of critical data, core to the business (e.g., customer, account) for stewardship and common management

Why Data Marts Proliferate: Business Semantics and Data Integration in Conflict

Robert Klopp

Director

Skyland Technologies

In most Business Intelligence implementations, data marts proliferate, creating a data management headache. Often this proliferation occurs despite the availability of a comprehensive, integrated, data warehouse.

The conflict between business semantics and data integration

- Is that integration requires a generalized, canonical, form... an Enterprise Esperanto that everyone speaks fluently
- While the reality of business semantics is that each end user constituency has, or will develop over time, a specialized language and several more specialized jargons
- So the natural linguistics of data semantics creates heterogeneity in the face of integration efforts to create homogeneity

The Role and ROI of Enterprise Schema Management

Peter Hallett

VP of Marketing

Schemalogic

Gartner says that “The average enterprise has a median of 14 databases and spends 60 to 70% of its application development creating ways to access disparate data.” Enterprise Schema Management is a new concept driven by the need to improve information sharing, cut costs and increase responsiveness to changing requirements. Involving repository technology and a highly collaborative process, success requires leadership from the IT and business architects who establish best practices. With increasing usage of XML Schema, now is the time to treat schemas as an enterprise asset to encourage re-use and ensure interoperability.

ONE of the Greatest justifications for ESM is RE-USE – increase programmer productivity, speed up the completion of projects. Additional justifications include:

- Requirement to share information
- Need to decrease costs/boost efficiency
- Need for greater responsiveness
- Requirement for better data quality
- Need to simplify integration & maintenance
- Need to reduce risk and disruption
- Demand to make current IT produce more
- The BEST REASON TO DO ESM is Consider what happens if you don't

Data Quality Assessment & Measurement: Developing Data Quality Process Measures

Shaun Williams

Data Integrity Manager

H-E-B Grocery Company

Joan Brooks

EDI/Data Integrity Program Owner

H-E-B Grocery Company

Ranked by Forbes Magazine as the 9th largest private employer, H-E-B Grocery Company has over 300 retail outlets in Texas and Mexico, 55,000 employees, and close to 10 billion dollars in annual revenues. H-E-B's systems process utilizes hundreds of information systems, many of which are legacy systems existing on antiquated platforms with little or no “on-line” data quality checks.

In 2001, the senior leadership team at H-E-B formed the Data Integrity Group with these primary objectives:

1. Assess the data quality issues in H-E-B's information systems, cleanse data issues, and help position the company for migration to more “modern day” applications.
2. Design and integrate data integrity as a way of life at H-E-B, through the use of data stewardship principles including process reviews, training, and measurements.

The Data Integrity department has spent the last 1½ years assessing and cleansing data that will be integrated with key Merchandising and Supply Chain applications. In addition, the group has

developed a methodology, process, and technology for tracking data quality exceptions within a given process, and assigning cost metrics to these exceptions.

This presentation discussed the processes, methodologies, tools and measurements implemented by H-E-B's Data Integrity department in order to improve and sustain data quality, by discussion of the following topics:

Design/Integrate Data Quality into HEB Culture

- Process, Job Design, Measurements, Training
- Data Quality is designed into business processes
- Processes/Partners are measured on Data Quality specific goals
- Data Quality responsibilities/accountabilities are designed into employee job descriptions
- Employees are trained that data is an important and key asset to HEB
- Data Stewardship
- Defined roles/accountabilities for managing business information and sustaining data quality

Metadata-Driven On-demand Data Integration

Patricia Klauer

CEO

Eclipse Data Systems, Inc.

Dina Bitton

Chief Technical Officer

Xtegra Corporation

This session described a metadata-driven data integration infrastructure that helps to preserve legacy code, while allowing the organization to move forward with new uses for the data. This approach not only solves present problems of data integration such as data migration and data mart proliferation, but also provides a platform for future growth and changing business requirements.

The authors addressed key issues such as:

- Persistently maintained data location and identification
- Standardized conceptual schemas
- Data transformation
- Preservation of local ownership and access controls

The Metadata repository allows creation of an enterprise information platform metadata store of enterprise data, regardless of OS, DBMS, location or format. It provides a global platform for standards enforcement and a horizontal data layer with unlimited reusability.

Extended business benefits include:

- Quick prototyping of new solutions & short development cycles for changing business needs
- Leveraging legacy/existing data dynamically
- Reduced maintenance costs, no new persistent data storage and data synchronization costs

Scientific Data - Challenges and Solutions

Olga Brazhnik

*Chief Database Architect
Synchronous Knowledge, Inc.*

Olga's talk covered:

1. The essence of scientific method;
2. The role of databases in the process of acquiring new knowledge
3. We base our everyday life decisions on the knowledge that comes from scientific research this is why it is important to know how it was obtained and how database community can improve its reliability
4. Scientific data are both abundant and extremely complex; they come from multiple sources that have variable reliability.
5. Models in science change constantly
6. Data tsunami in bio-medical science started with the ability of doctors to store raw data along with instrument readings and interpretations. A lab can easily produce terabytes of data per day the question is how much we need to really understand a phenomenon. We need a framework to organize our knowledge.
7. Practical solutions in science included data models, and reusable pieces of database design
8. Meta-models and various domain specific Markup Languages were discussed

Enterprise Semantic Models: Buy, Borrow, or Build?

Eli Israel

*Lead Modeler
Semantic World*

This session provided describe an introduction to semantic models and the roles they can play in an enterprise. It presented the qualities that identify a good model and the organizational factors affecting a model's adoption.

Additional questions answered include:

- Which industries already have mature semantic models?
- What types of organizations are best suited to adopting such models?
- Which team members should be part of a model adoption team?
- What is a realistic project plan for integrating such models into the business?

The relative strengths and weaknesses of using relational models, object models, and XML Schemas for this purpose were discussed.

Lessons Learned from Delivering Real-time Marketing Dialogs into Existing Operational Environments

Kevin Cavanaugh

VP of Technology

Unica Corporation

Effective marketing requires judicious use of customer data and coordinated treatment logic across many different operational systems and touchpoints. Unique data & business rule challenges are introduced when distributed operational systems are coupled with the analytical & marketing processing requirements needed to support real-time customer dialogs across these systems. This session explored system architectures and new data model requirements for coordinating outbound and inbound customer treatment strategies based on batch, real-time & event-triggered marketing scenarios across different touch points, including:

- Extended customer data model requirements to support cross-channel dialog marketing.
- Architecture, system design and data integration strategies to support both synchronous and asynchronous real-time marketing application demands
- Techniques for integrating & coordinating real-time, scheduled & event triggered marketing dialogs across touchpoints
- Important lessons from recent deployments of real-time marketing in high-volume production environments.
- Guidelines for establishing & prioritizing cross-channel dialogue based on business value metrics and implementation readiness/complexity

Standards-Based XML Management - an Insurance Case Study

Senthil Kumar

Insurance XML Manager

Fair Isaac Corporation

Insurance carriers spend millions of dollars each year processing policy applications and claims submitted by brokers and agents. This document-centric business is ideally placed to exploit XML, for which the ACORD standard has been defined for the exchange of policy information across the insurance supply-chain. ACORD addresses the challenge of standardizing extensions to documents to accommodate rapid changes in business.

This session focused on how one insurance carrier manages their XML documents with automated validation yet allowing for rapid development of extensions to their XML schemas, taking advantage of web services, business rules and predictive modeling.

ACORD has been adopted by:

- 72% of top U.S Property/Casualty Insurance Carriers
- 46% of top Life insurance Carriers
- 52% of top Global Re-insurance carriers
- 76% of Agency Management Systems

Savings:

- 20 -30 % of integration costs are saved annually by adoption of XML standards like ACORD
- 50% increase in process efficiencies
- Shortened Project development timelines
- Better Straight Through Processing Rate

Summary:

- Adoption of standards based XML Management can be the Magic Pill to relieve some of the pain associated with systems integration
- Significant improvement in internal process efficiencies (20%-50%). Some carriers reported efficiencies as high as 80%
- Overall improvement in integration efficiency (50%).
- Speeds up integration with partner systems
- High initial investment but significant ROI over a period of time

How to Uncover the Truth behind your Data

John Longley

*Data Cleansing Manager
Ministry of Defence*

Paul Nettle

*Data Cleansing Manager
Ministry of Defence*

This presentation describe how any enterprise can reap similar rewards by adopting the approach TAKEN BY The Cleansing Project (TCP) of the UK Ministry of Defence. These benefits were:

- \$30M savings generated
- Inventory reduced by 2.5%
- Supplier data reduced by two thirds
- Army, Navy & Air Force – coherent data
- Estimated \$60M benefits delivered

Key Lessons were:

- Get top level sponsorship
- Engage the dataset owners
- Team composition is critical
- Tool selection is crucial
- Use the experts
- Be innovative
- Always challenge – trust no one

Experiences with Meta-Data Management Across Tool Types

Christine Mandracchia

*Manager - Data Administration
American ReInsurance Company*

Effective meta-data management continues to be an undertaking requiring balance between the involved organizational roles and processes, and the capabilities of the specific tools available.

A combination of issues needed to be resolved in order to effect the "round trip" of meta-data among the data modeling tool, through a meta-data hub, to the ETL tool, to the BI reporting tool, and back:

- technical tool interface issues

- organizational roles and responsibilities
- work flow/process re-engineering
- scope and content management decisions.

Many of the work flow issues did not arise until after the technical tool interface issues had been resolved.

Meta-data policies that were adopted are:

- the business representatives now "own" the business data names, their definitions, and the approved corporate abbreviations and acronyms list
- there was a need to make provision for both "global" and "context-specific" business definitions
- the business definitions, business names, and their mappings to technical names will remain in the corporate data dictionary rather than in the data modeling tool and/or in the ETL tool (short-term)
- data structures originate in the data modeling tool, are imported into the meta-data hub, and then pushed to the ETL tool; data structures do not come into the ETL tool directly from the DBMS; the BI tool acquires the data structures directly from the DBMS
- most meta-data within scope is remaining in its "master" location, and in the meta-data hub, with little movement among the other tool types
- the types of meta-data within scope that need both a development and production version are the data models and the ETL job meta-data; the other meta-data within scope requires only one, production version

The SDLC methodology and project management protocols are being refined to allow the business representatives the opportunity to provide approved business names and definitions during the requirements gathering phase of a project, prior to or in conjunction with the data modeling efforts. Many of these work flow issues did not surface until after the technical tool interface issues had been resolved.

Open Source Data Warehousing and Databases

John Poole

*Distinguished Software Engineer
Hyperion Solutions Corporation*

The Open Source revolution is rapidly transforming the software industry in terms of both development practices and business models. Once regarded exclusively as the realm of software hobbyists, Open Source has become the software model of choice for many organizations. Data Warehousing and business intelligence (DW/BI) can benefit greatly from Open Source technologies, which enable both the construction of robust IT infrastructures, as well as an emerging class of software solutions for DW/BI, advanced analytics, and business performance management.

Motivations for using Open-source in DW/BI are:

- Advantages of building a DW/BI solution using Open-source tools:
- Enhanced ROI of the supply chain / corporate information factory
- Elimination of vendor-dependence and lock-in
- Management/integration of a complex supply chain

Schema Matching and Data Mapping Tools

Chito Jovellanos

President & CEO

forward look, inc.

This SIG session provided an overview of commercial products, in-house implementations, and R&D initiatives related to schema and data mapping tools. Mapping issues taken from the implementation of a data repository in the securities industry was used as a case-study. Several SIG participants indicated their interest in producing a survey paper on automated mapping tools and issues for potential presentation at EDF 2004.

Modeling Business Rules

David Hay

President

Essential Strategies, Inc.

Other than terms, facts, and certain constraints, data models are limited in their ability to portray business rule constraints. Indeed, the more generalized a model becomes, the less it is able to show the business rules that constrain the domain being modeled. Where the topic of the model is itself rules, however (as in regulatory agencies), then the rules themselves can be modeled.

In conclusion...an entity/relationship model fundamentally cannot show constraints on:

- Creation of occurrences
- Deletion of occurrences
- Legal values of attributes

Data structures can give form to rules, however, no matter how they are implemented

Service Based Architectures - Defining the Issues for Data Professionals

Robert Abate

Principal Partner & CTO

Intellisys, Inc.

This discussion reviews the components that have been found to provide a robust foundation for Enterprise Applications Integration [EAI], Enterprise Information Integration [EII] and Web Services. Specific implementations of the SBA/SOA were reviewed including issues and honest commentary on:

- Best Practices
- Standards (.net, Web Services, XML/ebXML, SOAP, J2EE, OASIS, Rosetta)
- Components & Topologies
- Service Layers & Content Integration
- Composites (observer/mediator)
- Wrappers & façades
- Business process alignment
- Patterns

Implementing an Integrated Enterprise-Wide Database for Applications

Ulka Rodgers

President

eTransitions, Inc

Project management lessons from a major integration effort were discussed. Management lessons learned were:

- Right people in right leadership
- Replace them if they don't fit
- Senior management buy-in
- Communicate UP as well as down

Project management lessons learned were:

- Clearly identify goals and objectives
- Tools & methodologies must fit culture
- Sufficient business analysts for project size
- Focus on the 80/20 rule
- Implement a process to ensure code quality (don't assume it)
- Testing time & resources cannot be compromised

Technical lessons learned were::

- Look for simple solutions for everything
- Simple solutions may not always exist
- Business rules are numerous, hidden, and complex
- Prototype every feature!
- Consider performance before presenting a UI layout or feature

Best Practices in Business Intelligence for ERP & CRM Suites

Aaron Zornes

Senior VP, Enterprise Analytics

META Group

Competitive differentiation depends on better analysis of ERP & CRM operations to drive continuous process improvement — & contribute to top- & bottom-line growth. This presentation gave META Group's bottom line prescription as:

- o Extend ERP & CRM vendors' DW/BI solutions
- o Leverage ETL & off-the-shelf data marts
- o Integrate near real-time web data feeds with traditional data warehouses
- o Supplement package-centric products with best-of-breed analytical applications
- o Examine & scrutinize vendor approaches to data warehousing — demand successful references

Federal Enterprise Architecture: The Business Process Models

Rob Cardwell

EVP

MetaMatrix

The Federal Enterprise Architecture (FEA) is a business-based framework constructed as a collection of interrelated "reference models" that will facilitate cross-agency analysis, and the identification of duplicative investments, gaps, and opportunities for collaboration among Federal agencies. The Reference Models act as a Target Architecture for which each agency will align. OMB and agencies will use the FEA for describing and analyzing information technology (IT) and other capital investments, and to improve Federal government service to the citizen. It includes a strong focus on delivering services to the citizen along with government- to- government process and information exchanges along with consolidating and integrating the services along lines of business.

The Business Reference Model serves as the foundation for the other reference models - the Data Reference Model, Performance Reference Model, Application-Capabilities Reference Model, and the Technical Reference Model. The Information and Data Reference Model (DRM) is an approach and a Federated Information Model that can be populated along government Business Lines and be used across Federal, State, Local and International e-government initiatives. The approach is based on both sound information and data base theory, a serious need, and an approach that correlates with standards organizations to create an open and extendable family of information models. These models can be one element of each organization's push for information integration and increased consistency, commonality, and visibility.

The DRM is:

- A framework to enable horizontal and vertical information sharing that is independent of agencies and supporting systems
- A framework that builds upon existing XML Schemas, Data Definition Libraries, and initiatives that exist across the Government (e.g., UN/CEFACT, UBL, ISO 11179)
- A collection of interrelated (or woven), business-driven XML Schemas
- A framework to enable agencies to build systems that leverage data from outside the immediate domain
- A repository that provides multiple levels of granularity to satisfy the re-use of data schema from multiple stakeholder views

Entities, Objects, and XML Schemas in One Model--Are You Crazy?

Richard Hecht

President

DATA Architects Technicians Analysts, Inc.

Is it possible that using different approaches in one modeling method can result in documenting and communicating data better than any one of these approaches by itself? This presentation explored a modeling methodology that combines the best of multiple approaches and furthermore maps the logical data to real physical implementations. This is not textbook and theory. This is actual practice and reality that arose from a need to help business users and developers better understand data and how it is implemented.

Specifically, this presentation addresses:

- Graphical techniques to display model information
- Approaches to present models to developers and users
- Concrete examples from real development projects

The examples showed how the synergism created from this novel approach helps produce schematics that improve the documentation and communication of enterprise data and its physical implementations. Applying these concepts and techniques can help produce meaningful and useful information to answer the questions developers and business users have about data and proves that combining multiple approaches may not be as crazy as it first appears.

Meeting the Requirements for Sarbanes Oxley

David Steinberg

*Solution Director
Idea Integration*

Peter Vink

*Business Development Manager
Idea Integration*

David Kotler

*Partner
Dechert LLP*

In response to recent corporate and accounting scandals, in July 2002, Congress adopted the Sarbanes-Oxley Act of 2002. In signing Sarbanes-Oxley, President Bush stated that its provisions are the “most far-reaching reforms of American business practices since the time of Franklin Delano Roosevelt.”

In basic terms, Sarbanes-Oxley created the Public Company Accounting Oversight Board and charged it with “oversee[ing] the audit of public companies that are subject to the securities laws, and related matters, in order to protect the interests of investors and further the public interest in the preparation of informative, accurate, and independent audit reports for companies the securities of which are sold to, and held by and for, public investors.”

At present, Sarbanes-Oxley’s requirements are directed only to “public companies that are subject to the securities laws.” Individuals within public companies – not just CEO’s and CFO’s – have new obligations and more severe liability (both criminal and civil) for improper acts.

What Does Sarbanes-Oxley Require?

- Requires CEOs and CFOs to certify financial statements.
- Requires rapid disclosure of material changes in the company’s financial condition.
- Requires reporting of all material off-balance sheet transactions.
- Requires rapid disclosure of insider transactions in company stock.
- Requires audit committees to be comprised of independent directors.
- Requires attorneys to report evidence of securities law violations to company’s chief legal counsel or CEO and, if necessary, to the audit committee or the board.

Getting Started

- Sarbane Oxley compliance approach should stem from your Accounting and Risk Management Control Processes.

- IT is an Enhancer and an Enabler.
- Work closely with Legal and Financial Management throughout the project.

What is a Control Process?

- It is a set of defined control objectives that may include policies, procedures, practices and organizational structures to reasonably ensure a specified goal is attained
- Typical control targets include effective and efficient operations, reliable financial reporting and compliance with laws and regulation

Deriving a Data Model for Service Architectures

Peter Aiken

Founding Director

Institute for Data Research

A metadata-based understanding is gained by a development process that applies eight transformations - organized into two phases - to each enterprise architecture/legacy system component. The eight transformations are applied in order to when effectively and efficiently developing an architectural component that is capable of delivering architectural and business engineering value. The transformations were illustrated in the presentation. Transformations and other forms of data analysis occur using model refinement and validation (MR/V) sessions in conjunction with key subject matter expertise (SME).

Supporting Corporate Data Integration Projects

Melvin Jones

ETL Architect and Project Lead

T. Rowe Price Investment Technologies

This presentation explored services, processes and procedures necessary to support data integration projects operating in a shared environment using repository-based ETL tools.

Conclusions and Recommendations

- Establish boundaries
 - Clear delineation of responsibilities between support and development
- Establish a data integration project support team
 - Full-time staff required when supporting multiple project teams
- Develop, maintain strategies, standards, processes, procedures
- Establish Roles and Responsibilities
 - Project team vs. Data Integration support team
 - Group and individual level
- Leverage existing IT processes, procedures and tools
 - E.g. Naming conventions, change control, versioning tool, vendor best practices
- Coordinate with other IT Organizations early and often
 - Avoids surprises
 - Maintain a separate lab to facilitate software evaluation and regression testing
 - Avoids conflicts with the operational environments
- Standardize on a single desktop install
- Package multiple applications for ease of maintenance.
- Create a timeline of events
 - Before, during, after

- Establishes clear direction and confidence that the upgrade will succeed
- Start with a Security Strategy
 - Protect ETL assets
- Implement Administrative Naming Conventions
 - Facilitates troubleshooting

SAP Business Information Warehouse "Top 10 Pitfalls" -- What Every DW Professional Must Know

Aaron Zornes

Senior VP, Enterprise Analytics

META Group

SAP's BW requires fine-tuning as indicated by other enterprises' "best practice" experiences. This will enable business to leverage its SAP R/3 investment to achieve maximum value. This presentation gave META Group's bottom line prescription as:

- SAP R/3 captures a tremendous amount of information which is often difficult to analyze
- SAP's BW is a very powerful set of enterprise analytic capabilities to access this info
- This requires ITO to optimize this investment via a solid understanding of technical infrastructure pain points

Department of Defense Net-Centric Data Strategy

Alan Perkins

Senior Solutions Architect

ASG Federal

In May of 2003, the DoD CIO announced a new Net-Centric Data Strategy that radically changes their data management and delivery paradigm. The transformation will be from standards-based, build-time data administration to consumer-based, run-time, network-centric data/information delivery. The vision is a virtual "marketplace" where data Producers and Consumers find each other and "trade" information Commodities. The approach has these tenets:

- Only handle information once
- Post before processing
- Consumers "pull" data when it is need in the form it is needed
- Consumers must be able to collaborate with experts
- Environment must be reliable and secure

The Goals are:

- Make data visible
- Make data accessible
- Enable data to be understandable
- Enable data to be trusted
- Support data interoperability
- Be responsive to consumer needs

The Critical Success Factors are:

- Communities of Interest (COI) -- to address organization and maintenance of data.

- Enterprise Architecture governance -- DoD and Federal
- Shared data spaces -- COI-managed mechanisms for data and metadata
- Global Information Grid (GIG) services -- for delivery
- Bandwidth enhancements -- improved infrastructure

In addition, there are a number of issues that, if not addressed, will result in sub-optimal implementation of the strategy:

- Governance requires a world-class, quality Enterprise Architecture
- Extensive teamwork, cooperation and collaboration is required within and among COI's
- Deployment strategy must be defined and implemented consistently
- Security of data and access security must be "bullet proof"
- Ensuring data and information quality is paramount
- Transition and modernization of legacy applications is necessary

In other words, the entire DoD data environment must be "net-centric."

3-D Knowledge Models

Henry Feinman

Information Architect

HJF Infosolve

The structure of Enterprise knowledge is complex – as complicated as any concrete product of our modern organizations. Other areas of endeavour have encountered the problem of complexity: Automotive design, structural architecture, computer chip design. In all these modelling domains the solution has been to use layers, perspective and 3-D CAD to build models that guide creation of these complex objects.

Firstly: Data – Structured and Unstructured, Business rules, Process, Entymologies – which of these belong to our modelling endeavours? What part of Enterprise knowledge is captured by our models, what part remains without?

Next: What alternatives have we for perspective and layering techniques: Data / Process split; Zachman's enterprise framework; the traditional conceptual / logical / physical; abstraction. What are the strengths and weaknesses of different diagramming methods – ER, UML, ORM?

This presentation stepped through the model using VRML, parts explosions, and other tools available to modern CAD modellers to see what insights become available using an integrated 3-D knowledge model. It answered the questions:

- What is Enterprise Knowledge?
- What are the strengths and weaknesses of current tools for modelling it?
- 3-D CAD tools – what can a third dimension add?
- VRML - what can the fourth dimension add?
- What insights become available to the integrated view of Enterprise Knowledge?

Tools for Modeling Privacy Requirements

Karen Lopez

Principal Consultant

InfoAdvisors, Inc.

Privacy is primarily a management problem, rather than a technical one.

Recommended Actions for coming to grips with privacy issues:

- Determine what efforts are currently underway at your organization
- Analyze past privacy violations
- Initiate a review of current data management deliverables to determine what personal information is being stored
- Develop privacy plan
- Perform Privacy Assessment on current systems (Automated and Manual)
- Update current Methodologies and policies to address privacy activities
- Ensure other corporate privacy initiatives are aware of IT assets for their use
- Train all staff with access to personal information on privacy fundamentals and corporate privacy policies
- Monitor and enforce privacy policies

Harnessing Technology to Fulfill Compliance Requirements

Jeff Canter

Vice President of Operations

Innovative Systems Inc.

Companies evaluating OFAC software should consider the significant costs of implementing the interfaces between the OFAC software product and the many administrative systems they have. In doing so, you soon realize that the OFAC software cost ceases to be a major factor when compared to the cost of implementing the interfaces. Consequently, it is critical to select the best system from the most stable vendor available to minimize the possibility of having to redo the interfaces later.”

Selection Recommendations:

- Identify an established provider with experience in compliance
- Select a tool that can be used for a variety of customer name-based compliance efforts, as well as customer data integration initiatives outside of compliance
- Ask if the tool prepares the compliance list – does it cleanse and standardize the entries to ensure effective, reliable matching
- Determine how list updates will be managed
- Keep the user community in mind – evaluate the software to determine if it is easy to use and maintain
- Look for products that will match on more than just name – including additional fields like address – to provide greater flexibility and accuracy in matching
- Understand how the system detects and corrects ‘false positives’
- Avoid the ‘black box’ tool syndrome – changing the business rules should be business user-friendly
- Find out if the product re-flags ‘safe list’ matches – matches that have already been reviewed and resolved

Managing the Downside of Data and Message Standardization

Chito Jovellanos

President & CEO

forward look, inc.

As part of the 'Risk and Pitfalls of Messaging and SOA' panel, this presentation highlighted commoditization, semantic monocultures, and complexity as issues to manage in order to offset the downsides of data and message standardization. Data metrics for scoping the impact of these factors were presented.

Theory vs. Reality When It Comes to Faster, Better, Cheaper Approaches for Information Integration

Dave Schrader

Director of Strategy and Marketing

Teradata, a division of NCR

Despite the hype, we still lack rigorous frameworks for evaluating the design and run-time impacts of new technologies and methodologies like Service Oriented Architectures, XML-based approaches to metadata and information exchange, and distributed vs. centralized approaches to enterprise information placement. This talk walked us through various architectures, explained how data warehousing solves many of the problems apparently motivating the latest hype cycle for EII, and outlines some experiments which are needed to understand not only the changes and potential advantages in design and reuse but also potential downsides in the run-time performance for various classes of applications.

Metadata Based Enterprise Information Integration

Arvind Shah

Managing Principal

Performance Development Corporation

1. True Process Integration cannot be accomplished without enterprise wide Data Sharing.
2. The Globalization demands On Demand Data Integration.
3. Warehousing has not been able to meet the changing needs of information integration.
4. Deployment of an Active Repository in conjunction with a Data Server Engine facilitates Real Time Data Integration of the data from disparate sources.
5. This Meta Data Based Layer between applications requesting data and physical data sources provides true program/data independence.
6. The resulting program/data independence also provides increased security.
7. All the stakeholders - IT as well as users, must participate in the Planning, Tool Selection and Implementation stages.

Customer Data Integration (CDI) Strategies for 2004

Aaron Zornes

*Senior VP, Enterprise Analytics
META Group*

Corporate master data is a critical asset that must be increasingly synchronized within & beyond the enterprise — primarily to solve business problems, not to re-engineer existing processes. This presentation gave META Group's bottom line prescription as:

- Apply real-time CDI engines to recognize customers instantly & treat them appropriately
- Focus on resolving political brinkmanship over customer data primacy (“who owns which customer data?”) as well as corporate data quality
- Add enterprise information integration (EII) solutions to ITO's 2003/04 shopping list as vendor attention to master data synchronization accelerates diversity & availability of offerings

Grids, The Semantic Web and Web Services: Finding Value In Three Visions of the Future

Zvi Schreiber

*Founder & CEO
Unicorn Solutions, Inc.*

These are three related visions which help to pool and abstract scarce resources:

- The Semantic Web may be applied to Corporate IT in the form of Semantic Information Management
- Semantics achieved by mapping physical schemas to an Information Model capturing the desired business world-view
- Capture formal business meaning of data at all stages of production not in retrospect on business reports
- Utilize to automate information management and information integration moving towards the real-time enterprise

Fast Modeling for Efficient Development: Prioritizing metadata by business significance (not syntax)

Joseph Maguire

Consultant, Author, Trainer

Session attendees learned how to accelerate the data modeling process without sacrificing the enterprise-wide view of data. Joe Maguire illustrated the following points:

1. Modeling notations categorize meta-data by syntax, rather than by business significance.
2. Modelers can quickly produce first-milestone data models for the development team, without sacrificing breadth and without resorting to vague, top-down modeling approaches.
3. Many constraints are about policy or about processing, rather than about fundamental definitions of data categories.

4. The syntactic artifacts that should appear on a first-milestone data model are limited to entities, attributes, identifiers, relationships, and maximum cardinality.
5. All other meta-data artifacts should be considered for inclusion on a first-milestone data model based on their business significance -- not their syntax. Those artifacts defining policy or processing needs can be excluded until later modeling milestones.

US Corporations enroute to Europe: Packing Data Quality can save Millions of Euros

Aaron Zornes

Senior VP, Enterprise Analytics

META Group

In 2003/04, data quality (e.g. cleansing, standardization, identification/matching, interdiction [compliance?]) will emerge from a disquieting enterprise embarrassment to a requisite core IT competency and budget item. Through 2005/06, data profiling technologies will become a conventional weapon for identifying data quality issues, planning data integration, and auditing information assets. This presentation gave META Group's bottom line prescription as:

- o Strive to improve internal awareness & communication about DQ issues & their impact
- o Quantify impact of DQ, suggest DQ-related data admin guidelines, & implement reusable DQ remediation logic
- o Monitor & remedy data closer to the source & implement business-level data management policies
- o Institutionalize DQ practices & match financial/material management rigor to information management
- o Don't rely on a "DQ czar" — for sustainability, DQ must be built into culture, IT, financials, & operations

Managing Data, Interface and Messaging Standards using XML Metadata Interchange (XMI)

John Friedrich

President and CTO

JRFII, Inc.

The industry is currently witnessing an important convergence of technologies, interoperability standards, and supporting tools. Specifically, the development and adoption of the Common Warehouse Metamodel (CWM), XML Metadata Interchange (XMI) format (both for CWM and Unified Modeling Language or UML) has fortuitously coincided with the great document, messaging, interface and presentation standards that have grown up around XML. This convergence presents an opportunity for leveraging the high and laudable level of industry support for these and related standards and technologies. This presentation described several examples, including:

- An XMI based repository
- Integrated environments for active management of data, interface and messaging standards
- How these integrate very tightly with existing commercial CASE, modeling, software engineering, data warehousing and business intelligence tools and repositories.

Introduction to Directories and LDAP

Bill Lewis

*Senior Technology Specialist
Cambridge Technology Partners*

Secure identity management is prerequisite to reliable user authentication and appropriate authorization of users to the digital assets of any enterprise. Directories have become the most widely-implemented single mechanism for storing data about individual and group identities. This session introduced participants to the fundamental concepts of Directories and the LDAP (Lightweight Directory Access Protocol) standard, and discuss requirements analysis, design, and implementation techniques for directory solutions.

How to Develop an Enterprise Data Strategy

Sid Adelman

*Principal
Sid Adelman Associates*

Today, if most CIOs are asked about the assets under their control (a primary asset being data), most would be forced to respond that the inventory of data is incomplete or nonexistent, little is known about the quality of data, and “there is no defined strategy for the productive use of this asset.”

This tutorial discussed how to go about developing an enterprise data strategy. It discussed how the key components of a data strategy can make a significant impact on the success of an organization, including:

- The components of the strategy
- The ROI of a data strategy
- The team to develop the strategy
- Selling the strategy to IT and to the business
- Categorization of Data

New Indexing Technology

Duncan Pauly

*Founder and CTO
CopperEye*

The past two decades has seen little innovation in data indexing technology. But new indexing technology has significant benefits; it increases system performance and longevity; delays hardware upgrades; provides competitive advantage through extra functionality; efficiently utilizes hardware resources and reduces development costs. This session discussed:

- What are indexes and why do they matter?
- What indexing mechanisms are currently available
- What new indexing techniques can provide
- How new indexing techniques impact application design.
- Comparative statistics on indexing performance.

Index Performance Determinants

- Key independence
- Size
- Distribution
- Cardinality
- Scalability
- Concurrency

Beyond War Stories: What can we Learn from Data Modeling Research?

Graeme Simson

Senior Fellow

University of Melbourne

Our body of knowledge is largely driven by "gurus" and some of our commonly held beliefs are not tested. Other controversies stand in need of resolution (need for routinely de-normalizing, level of implementation of subtypes, intrinsic stability of data were offered as examples).

There is good research evidence to support:

- E-R diagrams are a superior method of communication to ORM and relational models. Little research has yet been published on UML notation in this respect.
- Experienced data modelers operate quite differently to novices: they take a holistic picture and re-use patterns. Some important implications here for training and for valuing experience.
- Experienced data modelers develop diverse models which may reflect personal styles.

Graeme also discussed a number of broader issues about modeling:

- The need for modelers to have a sound understanding of implementation platforms;
- Poor modeling as a result of modelers focusing on user communication without corresponding attention to implementability.