CONFERENCE TRIP REPORT

14th Annual DAMA International Symposium
+
6th Annual Wilshire Meta-Data Conference

San Antonio Convention Center, San Antonio, Texas • April 28-May 2, 2002

Sponsored by DAMA International & Wilshire Conferences

This report compiled and edited by Tony Shaw, Wilshire Conferences

Contributing Trip Report Authors:
Many thanks for their contributions to this report go to the following individuals:

John Baughman
Paula Thornton
Elizabeth Gray

The DAMA International Symposium and Wilshire Meta-Data Conference were again co-located in 2002. The combined event drew an audience of over 900 attendees and speakers. To receive more information about this conference, and related future events, go to http://www.wilshireconferences.com

This report contains summaries of the key discussions and conclusions from many (not all) of the 100+ conference sessions, special interest groups, night school sessions, tutorials and workshops, as well as personal observations from attendees. Brief descriptions of the products and tools exhibited at the conference are also included.

Table of Contents:

Workshop

You Can't Cost-Justify Architecture!

John A. Zachman
President, Zachman International

Most people still think that the way to acquire funding for new systems is "cost-justification." John Zachman says this is no longer the case. The value proposition for systems has radically changed. Now Enterprise Architecture plays a central role in providing value to the Enterprise. There are four reasons why you "do" Architecture including alignment, integration, change management and reduced time to market. Without Architecture, there is NO WAY you can do any of these things.

Workshop

Developing Useful Use Cases: How to Avoid the "Useless Case" Phenomenon

Alec Sharp
Clariteq Systems Consulting Ltd.

Use cases offer great promise, but many analysts get disappointing results, because published methods are often overly complex, too focused on internal design, and ignore familiar techniques like workflow modeling and data modeling. Many of the common problems can be overcome by using a methodology that employs business events and three kinds of use cases:

- *business event:* object (entity) plus action (in "noun is verbed" format)
- *essential use case:* the *event* plus the specific actor handling or initiating it
- *use case:* an *essential use case* plus a particular deployment technology
- *use case scenario:* one or more *use cases* demonstrating a "vignette" with named actors, actual data values, and pre-determined decision outcomes

Other techniques are invaluable when developing all forms of use cases: a workflow model to establish the context for a use case, as well as the scope of use case scenarios, and a data model to establish terminology, definitions, and dependency. Data management professionals can demonstrate the value of (or need for!) well-structured, high quality data by showing its impact on a specific use case within the context of an important business process (as demonstrated in Amanda McLoone's excellent presentation).

Workshop

Universal Data Models to Jump-Start and Quality Assure Your Efforts

Len Silverston
President, Universal Data Models, LLC

Len Silverston provided a toolkit of the latest "Universal Data Models" for modeling people, organizations, contact management, product, and click-stream analysis. Sharing common modeling pitfalls and "war stories", Len illustrated the consequences of incorrect modeling. Among these were data anomalies, resulting from not distinguishing parties from party roles, from party relationships. Len showed how universal data models can be applied not only to data modeling jump-starts but also to data warehouse and data mart design jump-starts and quality assuring these efforts.

Workshop

Avoiding Catastrophe in Data Integration and ETL

Michael Scofield

Successful mapping of source data to target field depends upon a comprehensive understanding of the business meaning and data architectures of each source, and the target. It is not enough to use documentation (file descriptions, etc.) of sources (which may be obsolete). One must look at the actual data - all of it. There are step-by-step techniques for uncovering data anomalies, data quality problems, and semantical discontinuities in how a field is used. We start by creating

an inventory of the data, its architecture, and its behavior at each source, from the high-level view down to the specific, detailed behavior of each field and column, and inter-dependencies. Data profiling and domain studies are valuable tools at this stage.

The task of evaluating the commonality of any pair of source fields, and determining the appropriate target field in the target database is not for the naive. Mistakes are likely when "wimp analysts" make unwarranted assumptions about source data, without even looking at the actual business data (gasp!). In contrast, sound analytical techniques are required for getting the correct mapping and translation to the target database. Also, data quality issues such as validity, complete-ness, richness, and accuracy are critical. The speaker recommends an on-going data surveillance program to ensure that later production-ized loads of data will not be caught by surprise when a source changes definitions or scope of the data it supplies.

Workshop

OMG Model Driven Architectures - Modeling, Metadata, Middleware and Mappings for Enterprise Integration

Sridhar Iyengar
Distinguished Engineer, IBM Corporation
The OMG Model Driven Architecture (MDA) defines a new software architecture that allows integration and interoperability to be addressed across the application life cycle and not just between individual objects or components. MDA exploits the strengths of Modeling, Metadata, Middleware and Mappings – the 4 M's – into a unifying framework with UML and XML taking on a foundational role. The key standards and frameworks (some emerging) embraced by and integrated into MDA vision are:

- Modeling Standards (UML, MOF)
- Metadata Standards (XML, MOF, XMI)
- Middleware Standards and Web Services Middleware
- Component Frameworks
- Data Warehousing Frameworks (CWM)
- 'All-in-one/Uber Frameworks' (Microsoft .Net)

MDA allows a developer to design a model of an application or component once and automatically map these designs to several technologies. A key tenet of MDA includes reverse engineering that allows developers not familiar with modeling to incrementally gain the benefits of modeling and software architecture.

Workshop

Aligning Data Management Skills with Business Goals

Robert Seiner
Publisher / Principal Consultant, TDAN.com / CIBER, Inc.
What can be done to raise the value and recognition of the data manager? What skills do they need? It turns out that the skills necessary to manage structured data and unstructured data are very consistent. Managing content, intellectual assets and unstructured data can be tackled more effectively and efficiently when the strategists responsible for these activities leverage the experiences of the seasoned data management professional.

Bob Seiner discussed how to leverage what you know about data management to adapt to the quickly expanding fields of business intelligence/data warehousing, knowledge/content management and employee information portals. There are many similarities between data and unstructured data management as they relate to the implementation of stewardship & accountability for all information assets, the importance of building and enforcing data and unstructured data quality processes, and the impact that managing data and unstructured data will have on the organization culture. Bob also discussed the implementation of a meta data

management strategy that addresses the need for a catalog of enterprise artifacts that will become more valuable over time than the traditional meta data repository.

Night School

Common Data Architecture & ISO 11179 Comparison

Michael Brackett
Data Resource Design & Remodeling
The common data architecture is an integrated set of concepts, principles, and techniques that help people understand existing disparate data and transform those data into an enterprise-wide data resource that supports business activities. ISO 11179 is set of documents that form a standard for the development and maintenance of metadata that help people understand the existing data resource.

The common data architecture and ISO 11179 are in general agreement on most of the topics covered in the ISO documents. There are a few situations where ISO and the common data architecture are not in agreement. In these situations, the common data architecture is more robust and has a more solid foundation than ISO. The common data architecture encompasses ISO and goes well beyond ISO to cover all components of the data resource as well as existing disparate data.

Night School

The Process of Process Modeling

Thom Harrington
Systems Analyst, SAFECO Corporation
Process modeling is more than outlining the steps to accomplish a business process. This overview started with Zachman row one (scope) which uses Context Diagrams and creates a Business Process List**.** The Business Process Model (BPM) is a product of row two. It is developed using a four level approach. Levels one and two are high-level and represented in linear steps. Level three adds role or participant swimlanes and shows interactions between roles, decisions and loop backs. Level four adds notations for enablers and constraints. The BPM documents basic information needs and system update requirements. These can generally be referred to as Use Cases**.**

In row three, we can analysis the BPM by presenting information needs as Information Views (IV) and documenting the effect of system updates in Life History Diagrams (LHD)**.** IV's and LHD's can be summarized in the Entity/Event or CRUD (create, read, update, delete) Matrix. Inquiry and event processing can be shown in a Collaboration Diagram. When analysis is complete, results are interpreted in the Process Specification for understanding by developers.

Tutorial

Data Modeling – New Challenges, New Solutions

Graeme Simsion, University of Melbourne / Graham Witt, Tier Technologies
Graeme Simsion and Graham Witt looked at the role of the data modeler, and at techniques data modelers can use to increase the rigor of models and subsequent value to the business. Graeme compared data modeling to architecture, exploring the analogy in some depth. He argued that data modeling is a creative endeavor, and that the expert data modeler can lay foundations for applications that significant outperform those developed without professional data modeling input.

Graham Witt looked at a number of techniques for improving the quality of models, and supporting communication with business people. He received strong support for his views that XML does not remove the need for data modeling and that UML is a less than ideal tool for conventional data modeling. He also outlined a formal technique for presenting data models as "business assertions" for verification.

Tutorial

Applying Quality Principles To Data Definition And Data Modeling

Larry P. English
President, INFORMATION IMPACT International, Inc.
Larry challenged attendees to ask themselves — Are the knowledge workers you support:

1. Able to get the "kind" of information they need from their databases?
2. Able to understand clearly the meaning of the data and business rules?
3. More productive (effective and efficient) as a result of the information they get?

In support, Larry offered principles and processes to assess the quality of data definition and information architecture (data models and implemented database). Mr. English described quality principles from W. Edwards Deming's 14 Points of Quality, Masaaki Imai's Kaizen, and Quality Function Deployment and how they apply to data definition and data model quality. He described the business objectives of data modeling, and the business results of quality data definition and information architecture (data models and implemented databases).

The "products" of defined data, data models and databases are only "interim" products. Data definition and data models are the "information product specifications" for data and the databases are the "warehouses" for the data inventory. Mr. English described the 3 assessments to assure the quality of data definition and information architecture quality that make up the first process of TIQM (Total Information Quality Management). He offered concrete metrics to measure the costs of redundancy and metrics to measure the reuse of defined and implemented data, and to measure the costs of non-quality application projects with tangible dollars.

Mr. English then described how to perform TIQM Process 5 "Improve Information Process Quality" to increase customer satisfaction of the data definition and data model outputs. By applying Quality Function Deployment (QFD) you can design quality in to data definition and the data models and the resulting database designs (or evaluation of software packages). English concluded by outlining quality principles required to establish and sustain an information quality environment.

Night School

Introduction to Object Role Modeling

Terry Halpin
Program Manager, Database Modeling, Microsoft Corporation
This session introduced Object-Role Modeling (ORM) as a conceptual approach to information modeling, specifying data models in natural language that is formal and executable. ORM's graphic notation can capture more business rules than can be graphically depicted in ER and UML data modeling notations. Its attribute-free nature promotes semantic stability and facilitates validation by verbalization and the use of sample populations. Mining facts and rules from data use cases (central to the ORM discipline) can be adapted to ER and UML approaches. Since ORM models map in a fairly transparent way to UML and ER models, they can be used in conjunction with them if desired.

A fact-oriented approach to information analysis ORM offers benefits over other data modeling approaches by being designed to be:

- Understandable (facts and rules are expressible in natural language or intuitive graphics)
- Reliable (business rules are validated using natural language and sample populations)
- Capable (many business rules can be captured graphically)
- Stable (changes to the application domain have minimal impact on ORM models or queries)
- Executable (ORM's graphical language can be automatically mapped to DDL code)
- Using Microsoft's latest ORM tool, the following activities were illustrated:

- fact mining— going from data (syntax) to facts (semantics -- the meaning of the data)
- rule mining—determining constraints on fact populations, and rules for deriving new facts

One fact mining technique is to seed the data model with data use cases. Information can be presented in many different ways (tables, graphs, forms, maps etc.), and practice is needed in verbalizing these different visualizations to arrive at the underlying conceptual facts. The presentation showed a variety of examples to promote the required verbalization skills. It also showed how verbalization of facts and rules, in conjunction with positive and negative examples, can be optimally used to validate business rules, exploiting both positive and negative verbalizations of the rules. The different roles played by modeler and subject matter expert in the modeling process were also highlighted.

Night School

Overview of Dublin Core for Web-Based Meta Data

Todd Stephens
Director of the Metadata Services Group, BellSouth

The Dublin Core is a standard representation of metadata that can be used to represent semantic meaning for web based objects. A vast majority of the research articles looked at the documentation from a semantic point of view. Web pages contain an enormous amount of information designed for humans to read, but not for providing documentation or meaning. The semantic Web will provide structure to the meaningful content of web pages, which will create an environment where software agents can read and understand the content. Many researchers have expressed the concern that the web will not provide the vast potential promised until object representation is accessible to other machines or software. Searching, for example, would be more precise if web documents included the internal documentation required to express the actual content of the document.

SIG

Through the Portal to A Meta Data Community

Bob Carasik
Enterprise Architect, Wells Fargo Bank

About half the attendees reported that they were engaged in portal projects that had been successful. In both presentation and audience participation, key challenges faced included:

- Need to assign responsibility for official definition of taxonomy and dictionary terms.
- Some data could have multiple parties responsible.
- In the public sector there is need to share databases about resources.
- Desire to follow agreed standards is balanced by the cost of changing existing databases. Example: Environment Canada wants to follow FGDC metadata standards except where they are too voluminous.
- Meta tagging of generated pages is not useful, the taxonomy of the search engine must do the work for these types of resources.
- A UK government agency with 50,000 users reported that overly complex and deep taxonomy puts users off. The shallow taxonomy at Wells might work better for more general types of knowledge like those in the STEP portal.
- Comment from US Navy: channeling information to the right users, is a key issue, single sign-on is very helpful in implementing this.
- Enterprise meta models requires support and edicts from on high to get organizations to support them.

Night School

Requirements Gathering using Facilitative Techniques

Shelley Lieberman
Director of Strategic and Systems Planning, Mathtech, Inc

Requirements Analysis is a people-intensive process that identifies and analyzes the "what" of the business. There are multiple perspectives of the business, depending on who you speak to. What better way to obtain all those perspectives at once, and meld them into coherent business requirements than to hold facilitative sessions. Facilitation offers the opportunity to obtain requirements from multiple sources using a logical discussion path, with a controlled environment.

SIG

Standards for Data Naming

Judith Newton
Computer Specialist, Nat'l Institute of Standards and Technology
Methods of XML data transfer increases in popularity, yet the lack of consistent semantic content definition potentially poses significant problems. To avoid this Judith proposes applying the naming principles of ISO 11179, the standard for Metadata Registries, to XML namespaces may be part of the solution. A new project at NIST will address this issue. Anyone interested in contributing thoughts, suggestions, experiences, or other assistance is encouraged to contact Judith Newton at jnewton@nist.gov.

SIG

Data Resource Managements' Role in Data Warehousing

Amy Clark
Systems/Data Analyst, Blue Cross Blue Shield, CNY
To improve the development of a data warehouse, Amy emphasized the need to have a DRM team combining a variety of business backgrounds across the Organization with the skills/experience to perform the design and development tasks. She noted that the quality of deliverables are affected without an effective work request handling approach, because of the vast number of requests required to manage progress to completion. Amy discussed the value of providing Intranet access to the related meta data. Focusing on these considerations provide a faster delivery of a high quality product (the Data Warehouse).

KEYNOTE

The Implications of September 11 on IT

Ed Yourdon
CEO, Nodruoy, Inc.
Ed Yourdon kicked off the conference with a "mind-map" framework analyzing the strategic implications of the events of September 11. The major implications he identified were:

- Security – an obvious repercussion.
- Risk management – threats may now come from stateless persons or entities, with less predictable motives, thereby making risk management a far more complex task than previously.
- Emergent Systems – a new category of systems that assembles "on the fly", without centralized control or direction.
- Resilient Systems – able to withstand damage or security breaches without breaking. Analogous to biological systems, which degrade gradually under stress, rather than stopping entirely.
- Good Enough Systems – as the name implies, systems that are delivered because they serve a purpose adequately, if not perfectly, when speed to market is the highest priority.

Conference Session

The Same or Different: UML Class Diagrams & Entity Relationship Diagrams

Carrie Sherier
Data Architect, Williams Communications

Several Object Oriented terms have similar meanings, comparable to Data Modeling terms: class and entity; object and record; attributes; relationships (although UML relationships are more detailed); Generalization and super/sub typing; and multiplicity and cardinality. Other Object Oriented terms have no comparable data modeling terms such as operations, encapsulation, abstract classes and polymorphism. Carrie shared several common data modeling scenarios with corresponding class diagrams. She supported reasons to use both data models and UML models: different perspectives, different physical outcomes, and differing details about the subject. She challenged that all models should mirror the business in the same way and be compatible and complementary with each other.

Conference Session

From Entities to Stars, Snowflakes, Clusters, Constellations and Galaxies: A Methodology for Data Warehouse Design

Daniel Moody
Associate Professor, Norwegian University of Science and Technology
Senior Research Fellow, Monash University

This presentation described a method for developing dimensional models from traditional Entity Relationship models. The advantage of this is that it provides a more structured design procedure, which is based on the underlying relationships among the data. This allows data marts to be developed in an architected manner and simplifies extraction from production systems. The first step of the method involves identifying "event" entities in the data model – these correspond to fact tables in dimensional models. The second step involves identifying hierarchies in the model – these correspond to dimensions in dimensional models. The final step involves collapsing these hierarchies and aggregating transaction data to form dimensional models. A number of design alternatives are flat schema, terraced schema, star schema and snowflake schema. There is also a new type of schema called a star cluster schema. This is a restricted form of snowflake schema, which minimizes the number of tables while avoiding overlap between different dimensional hierarchies. Individual schemas can be collected together to form constellations or galaxies.

Conference Session

Data Stewardship: If only I knew then, what I know now

Carol Knight
Principal Consultant, Knight Consulting, LLC

Carol challenged that as data management professionals, we have been remiss in applying analysis tools, skills, and best practices to our own business issues. To implement successful data stewardship, we must clearly identify the problems we expect to solve and better appreciate the culture being impacted. There is no one data stewardship approach that can be implemented successfully in every environment. To develop data management functions (like data stewardship) that actually add value to constituents, we must become better analysts and communicators.

Conference Session

Managing Business Intelligence Programs - the Seven Streams Approach

Derek J Strauss
CEO, Gavroshe USA Inc.

Mr. Straus emphasised that Business Intelligence Initiatives should be seen as Programs NOT Projects. Such Programs should be ongoing, becoming more honed and sophisticated, as business needs change. To achieve this, and then to sustain it in a changing business environment, requires more than any "data-warehouse-in-a-box" or "30-day wonder cure" solution will be able to provide. One needs a Proven BI Planning Framework (i.e. The Seven Streams). Derek Strauss talked through his Seven Streams Framework :

- Corporate Data Model (Subject by Subject)

- Corporate Knowledge Co-ordination (Artifact by Artifact)
- Corporate Information Factory Development (Topic by Topic)
- Data Profiling and Mapping (Source by Source)
- Data Cleansing (Attribute by Attribute)
- Infrastructure Management (Component by Component)
- Data Quality Management (Element by Element).

Conference Session

Meta Data Partnerships: The Key to Populating the Dictionary.

Stan Slossberg
Director, Metadata Management, CIGNA
Stan Slossberg presented an approach deployed by CIGNA to ensure that metadata is collected and made available through a central repository. After developing a metadata strategy, getting the right tool and establishing best practices, metadata might still be captured in a territorial fashion, or not defined at all. A component of the overall metadata management approach must include a methodology for becoming a virtual project team member, and demonstrating the value of properly stored and reportable metadata.

Stan discussed how his Metadata Management team focuses on Partnerships and Metadata Mining, to ensure that Metadata inventory requirements are integrated into upfront planning activities, and that the specific deliverables, expectations, timeframes and roles/responsibilities are clearly established.

Rather than forcing application teams to provide definitions, the CIGNA Metadata Management team assist teams who have mission-critical deliverables that otherwise may exclude defining their metadata. The proof of CIGNA's success is a metadata repository representing the full enterprise. With customers knocking on the Metadata Team's door, the value-added is self-evident. Metadata Partnering and Metadata Mining have been key to the robust metadata implementation: weekly stats show an average of 16,000 site hits on the metadata repository website by over 450 unique visitors.

Conference Session

Common Warehouse Metamodel (CWM): An Introduction to the Standard for Data Warehouse Integration

John Poole, Hyperion Solutions Corporation / Dan Chang, IBM / Liz McIntosh, SAS Institute / Doug Tolbert, Unisys
In a panel format, members from the Common Warehouse Metamodel (CWM) community provided an overview of the CWM standard, including CWM's value proposition, architecture and technical underpinnings, and examples of using CWM to interchange an actual relational model. One strong message was that CWM enhances meta data return-on-investment by minimizing the costs of integrating multi-vendor, best-of-breed tools in the implementation of a data warehouse or information supply chain. The CWM representatives also provided live demonstrations of CWM-enabled products as part of the "CWM Enablement Showcase" in the conference vendor area.

Conference_Session

Model "Status" for DW/BI, Business Rules and CRM Success

Henry Feinman
Data Architect, HJF Information Solutions
"Status" is a cornerstone of business relationships and yet is rarely understood or modelled correctly. Correctly modeling Status enables the capture of accurate database history and derivation of data mart measures and facts from situations where none seem apparent. Client and account status are natural business concepts that are seldom modelled effectively. The reasons for this are varied, but can be overcome at the start – saving effort and resources.

Correctly modelled status gives us new, critical information for our Decision Support environment. It ensures accuracy of classifications needed for day to day dealing with client relationships, enhancing effectiveness in both CRM and Business Rules contexts.

Conference Session

The Selling and Re-selling and Re-selling of Information Management

Larry Dziedzic
Senior Information Management Architect, Johnson & Johnson
This presentation focused on the idea that the work to sell (promote) the functions of Information Management (particularly Data Management) never cease. It incorporated some the ideas and tools (software and games)that Larry Dziedzic had used over the years to both establish and re-establish Information Management in the eyes and mind of his customers and with all levels of management. The presentation provides some practical ideas and methods that can be utilized to effect the selling and re-selling of Information Management.

Conference Session

Data Standards Implementation - And You Thought Standards Development Was Hard!

Sara Hisel-McCoy
Data Standards Team, U.S. Environmental Protection Agency
Years after developing an ISO/ANSI 11179 metadata registry, the Environmental Data Registry (EDR), to serve as the backbone to the Agency standards setting process, the Environmental Protection Agency is now tackling the challenges of standards implementation. Having successfully established a core set of Agency standards, and with new standards in progress, the Data Standards Team is conducting outreach and education activities to inform program system managers on standards implementation concepts such as stewardship, conformance, and data harmonization. Ongoing challenges include maintaining management commitment, policy development, approval processes, and standards integration.

Conference Session

How Data Administrators Can Survive the XML Revolution

Charles Dietz
Director, Data Administration, MetLife
The focus of Data Administrators will have to shift from fixed back-end data stores to flexible front-end message based systems. But data administrators can use their existing data models and metadata to create a model-based XML internal standard and an environment to manage and control multiple XML vocabularies linked to back-end systems.

XML is an important topic for Data Administrators because, in the future:

- Virtually all B2B communication will be conducted via standards based XML transactions
- The primary interface to large data stores will be XML, either converted to or from database formats or stored as XML
- Traditional data modeling will be replaced by a hybrid of XML Schema development along with mappings to back-end system data stores

Conference Session

Meta Data Lies: Data Profiling Is Your Lie Detector

John Howe
President,J.W.Howe LLC. Management Consulting
The main weakness in traditional approaches to executing data migration or data integration projects is that they all assume that metadata is accurate. In reality, most metadata is not accurate. Most of it is out of date. Data Access, Profiling and Mapping creates a factual body of

knowledge about your data. This knowledge will provide significant reductions in the risks and costs associated with accessing, moving, cleansing and transforming data in any type of project.

Conference Session

Data Model Management: Keeping the Logical and "Real World" in Synch

Ralph Mohr
Director, Data Warehouse Architect, Covansys
A model management strategy provides the processes and procedures required to ensure the maximum return from an organization's data architecture investment. The strategy should address:

- Model types and usage
- Change control management
- Model standards
- Metadata
- Roles and responsibilities
- Model repositories

Conference Session

Data Management at the Project Level

Thomas Zaborsky
Data Analyst, CNA Insurance
The premise of Mr. Zaborsky's presentation is that the vision of Data Resource Management (DRM) can be grown within an enterprise by actively practicing DRM disciplines at a project level. It's at this level that DRM can more readily justify its existence and provide tangible values to both the business enterprise and the systems organization. Attempts to try and introduce DRM concepts across an entire corporate entity will often fail. Conversely, it can often succeed when focused upon an individual business unit. This takes a strong time commitment from persons involved in DRM to stay committed to their organization in order to make the DRM vision thrive.

Conference Session

Globalized Data for the Web

James Bean
President and CEO, Relational Logistics Group
The web is "borderless". By default, all web applications are exposed to the international community as potential markets. To effectively capture these markets and conduct "global e-commerce", there is a significant dependency on international and industry data standards. The flexible and extensible metadata capabilities of XML can be leveraged to support the definition of global e-commerce transactions, as well as acting as an enabler for Global Customer Relationship Management (G-CRM) and Enterprise Application Integration (EAI).

Conference Session

What's So Spatial About Spatial Data? Integrating Geospatial Data into the Enterprise Data Resource

Michael D. Walls
Software Engineering Manager, PlanGraphics, Inc.
Geographic Information Systems (GIS) technology evolved separately from other information management technologies. This was fine when GIS concentrated on computer cartography and left management of nongraphic attributes to mainstream DBMS technologies. Now, however, improved GIS and RDBMS data handling technologies and increasingly urgent business needs are forcing a convergence in these two aspects of data management.

The fundamental challenge is that geospatial data within a GIS adds whole new layers of complexity in addition to those data administrators are accustomed to dealing with when incorporating more traditional data types within the DBMS. Even the inclusion of imaged documents or raster photography could be accommodated relatively painlessly, once we had data types like BLOBs to handle their unique storage needs and raster viewing tools available for their display. In contrast, GIS features have geometry and explicit locational coordinates expressed in specialized and quite technical units of measure. They often have topological characteristics and relationships that further extend the bounds of data management.

Conference Session

Data Quality and the Importance of Meta Data at Centrelink - A Case Study

Peter Davis
Manager, Data Management, Centrelink
Centrelink is a large government agency which makes payments on behalf of many client agencies. Its complex and rapidly changing processing environment is supported by a central repository and code generation facilities – all relying on the system metadata . The use of code generation for file definitions, database updates, screens and screen field validation, XML, help and error messages enables a high degree of agility and accuracy for its processing environment.

Peter outlined Centrelink's major future challenges, including supporting multiple service delivery channels across multiple platforms and managing multiple repositories to do it.

Conference Session

To Laugh or to Cry? Persistent Prevalent Database Fallacies

Fabian Pascal
Analyst, Editor & Publisher, DATABASE DEBUNKINGS
A lot of what is being said, written, or done in the database management field – or whatever is left of it – by vendors, the trade press and "experts" is increasingly confused, irrelevant, misleading, and outrightly wrong. While true of computing in general, in the database field the problems are so acute that its practice and technology are regressing/ Fabian debunked the following industry misconceptions with entertaining quotes from notables and 'knowables':

- If you need "complex"data types", you cannot use a RDBMSs and must use an ODBMS
- You don't need to know business rules to design database tables
- If you want good performance, denormalize
- If you use a RDBMS with object applications, you must enforce integrity in the applications
- RDBMSs store disassembled objects and force you to reassemble them each time you need to access them

KEYNOTE

Trends in Metadata/Data Management

Peter Aiken
Founding Director, Institute for Data Research, Virginia Commonwealth University
Several surveys of world wide data management practices indicate several interesting conclusions:

1. A survey of self reported scores indicates that most organizations are not up to an equivalent of level III Data Management Maturity - this includes data and metadata management practices

2. We are beginning to establish that lack of maturity of data management has resulted in project cost over runs, schedule delays, and systems implemented with less than full functionality.

3. There exists quite a bit of disconnect between the expectations of organizational IT management and data managers regarding the promise of XML and in particular, most organizations are not seeking to take advantage of the potential organizational savings that XML offers.

More needs to be done to educate management as to how to utilize this untapped potential and realize immediate IT savings.

Conference_Session

The Politics of Data Analysis: Working Within the Constraints of Corporate Data

Amanda McLoone
Business Process Engineering Manager, Intel Corporation
Many corporations are recognizing the negative impact of poor information quality and are investing in improvement. This should create jubilation since historically, data analysis, has been perceived as low value, especially on time-constrained projects. Excitement wears off quickly, though, as it is apparent that in large corporations, poor information quality has roots so deep, that in spite of best intentions, most barriers remain in tact. Regardless of the root cause of poor data management practices, the need to appropriately apply theory and practicality is critical to successful data analysis.

Using a case study from a Fortune 500 manufacturing company, this presentation identified common constraints encountered in corporate environments. The presentation did not advocate eliminating the use of Information Quality theory, but rather to recognize the conflict between theoretical practices and practical application in large corporate environments.

Common Corporate Constraints Addressed:

- The Corporate Data Foundation (Product, Customer, Order) -- Constraint: Fixed Structure Legacy Data
- The Enterprise Data Warehouse -- Constraint: Invisible Data Transformation
- The Enterprise Application -- Constraint: Proprietary Data and Data Structures
- The Programmer -- Constraint: Unstructured Data
- The Organizational Silo -- Constraint: Unintegratable Data

Conference_Session

Building a Bridge Over Disparate Waters

Sandra Hostetter
Manager, Content Management Group, Rohm and Haas Company
The corporate world is drowning in disparate data. Data elements, a.k.a. field names, column names, row names, labels, metatags, etc. seem to reproduce at whim. Librarians have been battling data disparity for over a century with tools like controlled vocabularies and classification schemes. Data Administrators have been waging their own war using data dictionaries and standards. Both camps have had limited success with their strategies. Why? (1) These tools are not a common data architecture and in some ways actually contribute to the problem. (2) Librarians and Data Administrators are not working together to find a total solution.

Conference Session

Resolving Context and Meaning in Computer-Based Communications

Chito Jovellanos
President and CEO, forward look, inc.
Using examples from the securities industry, Chito detailed a network-enabled service to mitigate reconciliation effort and costs in transaction-based environments. The approach relies on information theory and multi-variate statistics to quantify relative variances in content and meaning between transactions exchanged across counter-parties. This technique facilitates straight-through-processing (STP) by enabling the sender's transaction to be unambiguously understood by the receiver. The physical implementation was described as a lightweight array of nodes on a local area network, where each node represents a topic-aware agent that resolves the meaning of each sub-component of a transaction.

Conference Session

Enterprise Data Integration: Development of an Enterprise Data Model

Noreen Kendle
Enterprise Architect, Delta Technology - Delta Air Lines
This presentation was a real life success story on the creation of an Enterprise Data Model for Delta Air Lines. It described a practical and applicable six step iterative approach balancing the strategic view of "top down" with the reality of "bottom up". The Enterprise Data Model is being used to create the Operational Enterprise Data Stores, integrating data across the airline. The Enterprise Data Stores play a major role in Delta's Smithsonian awarded DNS (Delta Nervous System).

Conference Session

Implementing Information Stewardship: Data Definition & Beyond

C. Lwanga Yonke
Aera Energy LLC
Aera's Information Stewardship Program is structured around seven major stewardship roles:

1. **Information Producer:** accountable for integrity of data content
2. **Data Loader:** accountable for integrity of transcription of data from one form to another (e.g., from paper to electronic). In some instances, accountable for inspecting and correcting data prior to loading into an electronic database
3. **Knowledge Worker:** accountable for integrity of data usage
4. **Manager or Supervisor:** accountable for the quality of the data created, loaded within their unit, and the responsible use of that data
5. **Information & Process Owner:** accountable for integrity of the definition of a business process or business value chain, along with the relevant data requirements
6. **Technical Support Staff:** accountable for the quality of the information resources and technical infrastructure, and for the proper deployment, training and support of projects that meet the information requirements of the business
7. **Data Definer or Business Information Steward:** accountable for the definition of data elements, business rules, data quality levels, etc.

Thirty individuals were assigned as Business Information Stewards for the 53 subject areas identified in the Conceptual Data Model. In addition, Information & Process Owners were identified for all the major business process. The remaining five roles are being implemented as follows:

1) Identify the data of interest
2) Identify the data value stream and the relevant work processes

3) Identify the stewardship roles for that value stream
4) Establish role-specific Data Quality expectations & accountabilities
5) Map each stewardship role to specific individuals
6) Communicate expectations & accountabilities, and train all the players
7) Establish measures & report data quality levels
8) Be available to help

Effective change leadership was key to the success of the stewardship program. Significant business results have already been delivered since the initial implementation: a) data quality is increasing, b) time spent finding, correcting and reconciling bad data is decreasing, and c) data reuse is improving, reducing the cost and cycle time of IT development.

Conference Session

Meta Data Success = Service Based Organization

Todd Stephens
Director of the Metadata Services Group, BellSouth
The Metadata Services Group within BellSouth has spent the last 2 years developing Metadata based products. However, the successful collection of metadata is only half the battle. Organizations that develop metadata solutions must begin to transform themselves into a services based organization. Failure to focus on services will result in another failed metadata project.

Conference Session

Data Management's Next Big Thing

Len Silverston, Karen Lopez, Robert Seiner, Graeme Simsion, Tony Shaw (moderator)
This fast-moving panel session discussed numerous trends in data management that may have significant implications for the future:

- Unstructured Data -- Bob Seiner, editor of TDAN.com, predicts that the need to manage unstructured data, including text, geospatial data, video, audio, etc. is a huge trend, especially as it relates to content and knowledge management. Traditional data management practitioners should watch this trend closely, because they have the right skills to contribute significantly to this emerging field.
- Information Privacy – Karen Lopez of InfoAdvisors highlighted information privacy requirements as on of the next big things for data management professionals. As many of the Fair Information Practices require meta data management and an enterprise view of data, she believes that data architects are best prepared to step up to the challenge to assist their organizations to meet these requirements. Data architects who ignore these trends towards more personal data privacy may find themselves unable to meet business demands to meet customer demands.
- People Integration – Len Silverston of Universal Data Models Inc. predicts that the human component of IT implementation needs much more attention than it has in the past.
- The Death of Data Management – Graeme Simsion provoked the audience with the assertion that centralized data management does not work. He suggests that the traditional notion of a highly “controlled” data function is mostly perceived by the rest of the business as “getting in the way.”
- Outsourced Data Management – none of the panelists felt that outsourcing of a strategic function such as data was a good idea.
- Data on the Balance Sheet – the audience asked when the “value of data” would appear on the corporate balance sheet as an asset. The panels view was that, in the wake of recent accounting scandals, it would take a long time before accountants would settle on standards for valuing data.

Conference Session

Preparing for CRM: What Data Managers Should Know

Jill Dyche
Vice President and Partner, Baseline Consulting Group
Jill introduced emerging issues around data management for Customer Relationship Management programs. With almost every company is either considering, planning, or implementing CRM, Jill focused on the urgency of incorporating data as a staple in all stages of CRM development. (She maintains that in the rush to embrace "People, Process, and Technology" data is often left out of the equation altogether.) Jill reviewed some of the surprises companies are currently confronting during CRM – particularly when it comes to integrating data from disparate sources. She insisted that a formal data management function – separate from DBAs and database designers – is no longer a luxury for many companies, but a critical factor for initial and ongoing CRM success.

Conference Session

Data Model Quality: What Is It?

David Hay
President, Essential Strategies, Inc.
After twenty-five years in development, data models (and their cousins, Object Models) appear to have come into their own in our industry. The problem is that there are many different approaches and many different attitudes toward data modeling. It is becoming clear, however, that some characteristics of data models are better than others. The first premise is that the quality of a data model is directly proportional to its ability to communicate concepts to non-data modelers. The data model is a communication tool to establish that an analyst's understanding of the nature of an enterprise is in fact correct. If it cannot do that, it is not worth the effort. Data models can be evaluated in terms of their graphics and the way they are presented.

Conference Session

How to Estimate Data Modeling Project Efforts

Gary Flye
Manager, AVP, Wachovia Corporation
Data modeling project estimates need to be fast, accurate and flexible to provide answers to cost, time and resources. Gary offered an automated tool using Microsoft Excel that uses metrics gathered from past modeling projects to estimate cost, time and resources for any new modeling Project:

- General project facts (number of entities, attributes, relationships, tables, columns and keys)
- The skill level of the modeler(s)
- The amount of analysis that was needed
- Number of hours spent on logical and physical design.

Combining results across projects, ratios can be developed to provide reliable estimates for future projects. Gary cautioned that this kind of tool should never be a replacement for good judgment!

Conference Session

Meta Data Integration Tools

Attila Finta
Director, &
Marcia Rhode
Director, A.M. Consulting, Inc.
Five tools integrating meta data from heterogeneous sources were examined.

The two ETL-centric tools – DataStage XE with MetaStage by *Ascential,* and PowerCenter with Metadata Exchange SDK by *Informatica* – were defined as 'suitable' in those environment that already have the tools and are mainly doing meta data integration with data warehouse sources, and you need the strong orientation around reuse that these tools offer.

Another ETL environment oriented tool – MetaCenter by *Data Advantage Group* – was defined as 'suitable' in Informatica and Business Objects environments needing a customizable business-friendly interface to meta data.

The two enterprise repository solutions – Rochade Repository by *ASG,* and Advantage Repository for Distributed Systems by *Computer Associates* – were defined as 'suitable' when meta data integration needs are not data warehousing centered, include rationalizing many diverse meta data sources (especially systems development tools), and can make use of the powerful versioning and extensibility capabilities these tools offer.

Conference Session

Managing Information Quality in an Integrated World

Danette McGilvray
Agilent Information Quality Program Manager, Agilent Technologies
Poor data quality — the cause of many problems in any data migration project. Danette shared the approaches, tools, and processes used to improve information quality (IQ) in a large-scale data migration project at Agilent Technologies. In migrating all ERP (Enterprise Resource Planning) systems worldwide - finance, procurement, order management, manufacturing and planning - Agilent is focusing on the front end (fixing data prior to migration) to save five times backend costs due to rework, delayed schedules, and data that will not load.

Specific recommendations included:

- Implementing tools and processes for data analysis in the source system as early in the migration process as possible. Comparing the source data to the target requirements identifies gaps. Options for tools and process flows were shared.
- Data cleansing in the source systems is always the first 'line of defense', focusing on: 1) current business requirements sans new target system; 2) addressing 'gaps' previously noted, and 3) changes identified via results from testing loads to the new application.
- Where legacy system changes are not possible, the same specifications are applied to the transform and load programs. Be prepared to create data required by the new systems that doesn't exist in legacy systems or replace data with quality so poor that it cannot be used.
- Team structure and roles and responsibilities were covered, including relationships with teams outside of the data migration group.
- Proactively addressing IQ resulted in ahead of or on-schedule test cycles and an overall increase in successful record loads.
-

SIG

Staffing the Data Organization

Jane Carbone
Partner & Co-founder, &
MaryAnne Reuther
Partner & Co-founder, Infomajic, LLC

While organizations may have a keen interest in planning and implementing the data changes they need to make, they may be unclear about what is required to position an effective data team. Jane and MaryAnne suggested that a good first step would be to examine current staffing practices (e.g., how data team members are selected; what outputs they are responsible for). Since the data team plays such a key role in enabling business goals, staffing criteria needs to include not only technical competence, but other skills and knowledge, management

competencies and breadth of experience required for success. Key roles to consider filling include: Information Steward, Chief/integration Architect, and Data Planner. Managers also need to assess the organization "culture," e.g., readiness to change and acceptable level of centralization, when they want to move from a "Business as usual" organization structure to one that enables an "Information is our business" position.

SIG

Meta Data-Driven Content Sensitive Security Management

David Schlesinger
Data Entitlement Security Program Manager, Intel Corp.
Mr. Schlessinger proposed how use of content descriptive meta data in a central meta data repository allows access security automation. Three breakthroughs used at Intel Corporation: (1) Define each user role in applications via meta data (documenting data sensitivity and related security policies for data use); (2) Require that application owners define security rules related to data and specific role use; (3) Separate initial entitlement process from daily access security; automate the entitlement based upon the metadata using the security rules for access privileges and the worker's HR attributes (daily HR data validation removes entitlement to specific user roles if worker no longer qualifies).

Conference Session

The Use of Assertions in Information Modeling

Graham Witt
Senior Consultant, Tier Technologies (Australia)
This session described a modeling technique and associated toolset which addresses the fact that reviews of an information model by business stakeholders, subject matter experts, etc. often fail to identify flaws in the model. This is because models are most often presented in a diagrammatic manner which is fine for a high-level overview but deficient in two respects:

1. lack of detail
2. the question "have we covered everything?" is not always easy to answer.

By contrast, this technique generates a set of assertions about the model which can be reviewed individually, each reviewer disagreeing, agreeing or seeking clarification. Once an agreed assertion set has been achieved (by changing the underlying model), the model can then be used to generate a database design. Information modelling has two audiences: business stakeholders and database designers. Diagrammatic techniques do not adequately serve to assure effective reviews of information models, so that many systems are built or acquired which do not adequately meet business needs.

Conference Session

Corporate Data Categories

Jeremey Janzen
Data Administrator, British Columbia Ministry of Forests
In any organization, "data is defined, collected, transformed, used, summarized, and reported for the purpose of making business decisions. Data defined and collected by one person in an adhoc manner, with no standard agreed method is cheap to get, but of far less value than data collected for a defined business purpose. Locally defined data (office or department, for example) while quick and effective initially, becomes costly as it is shared with others. Adding appropriate context requires translation, reformatting, explaining, redesigning, converting, and copying. These ancillary costs may outweigh what appeared to be quick benefits before.

Conference Session

"Data Driven" Enterprise Architecture

Richard Hecht
President, DATA Architects Technicians Analysts, Inc.
No matter how much we recognize the need for and importance of an architecture, it has to be used and provide value. Richard described how architecture products have been used to support current IT strategic initiatives and how they have provided value to organizations. Central to his message was using data to create a centralized, easy-to-use knowledgebase. He also emphasized focusing on data as the basis for blueprints to provide the context information presentation — making it meaningful to both users and management.

Conference Session

XML: Is it a Better Implementation of Relational Theory than SQL?

Mark Milodragovich
Senior Information Engineer, Nimble Technology
In a presentation that could have been entitled "Back to the Future", Mark Milodragovich described the convergence of XML with Codd's early Relational Theory. XQuery (the new XML query language) and RDF (the semantic layer of "The Semantic Web") share the same mathematical foundation as the Relational Calculus. Twenty years of SQL pros and cons can serve as a history lesson if we examine our practices closely. Now, the XML tool/vendor scramble is on and the DRM community needs to be a driving force with a clear goal. Milodragovich argues that our goal can and should be a network-based implemention of the original relational theory.

Conference Session

A New Approach to Data Standardization - One that Works

Harry Ellis
Senior Consultant, Army Data Services, The British Army
Harry addressed the problems presented by the incompatiblity of data definitions across many databases, typical of a large and complex enterprise. He suggests that the languages used for modeling often contribute to the problem and offered a new language for conceptual modeling. Developed by the British Army and used with encouraging results, Harry shared the concepts and strengths of language.

Among them was an explicit notation for the categorization schemes and type designation schemes that are found in the real world but are hard to recognize in an entity or UML model. He also suggested that data modeling moves through five levels beginning with the description of extant data and the capture of specific business needs. These are developed through a rule-based process of technical refinement to create compatible models that can be harmonized in the best interests of the enterprise overall. Go to www.cbml.co.uk for additional information.

Conference Session

Data Strategy: Global Modeling & Knowledge Management for Local Content

Stephan Stadelmann
Partner, FINSTRAT AG
Stephan provided an overview on the challenges & benefits in designing data architectures and streamlining data management in a global setting. Based on an implemented data strategy in the Asian HQ of a leading information provider, it was demonstrated that data management can contribute substantially to the growth and effectiveness of business at both a domestic and global level. Solutions to key challenges such as language, reference data, multi feeds and legacy systems in constantly changing environments were also discussed.

Conference Session

Partnering for Static Data Management...the Foundation for Business Rules

Barb Morgan
Vice President, Bank One, &
Catherine Nolan
Enterprise Codes Analyst, Allstate Insurance Company

Efficient management of meta data (including codes) requires a different underlying structure than applications need to drive their editing, decoding and translation activity. Populating application tables with valid combinations of codes allows the constraining business rules to exist outside application code and lends itself to the timely implementation of change.

Allstate developed an efficient centralized meta data change management process to provide the highest quality and consistency across applications. Success relies on close, continuous cooperation from all participants (business, data management, database administration, and applications development). Packaged solutions can be driven by the same centrally managed business rules constraints if architected to drive edits from user-defined codes, sourced from data feeds external to the package (that is, the meta data management environment).

Conference Session

Data Modeling and Software Engineering

Karen Lopez
Principal Consultant, InfoAdvisors, Inc.
Karen gave an overview of the Software Engineering Body of Knowledge (SWEBOK), including highlighting where information architectures are addressed. She pointed out that much of the data-related discipline is not covered within the SWEBOK. She also discussed the efforts to implement a professional licensing scheme in the U.S. and Canada, as well as identified reasons why the IT profession is not currently mature enough to support a licensing program. She also outlined the risks to professionals, organizations and the public in rushing to a licensing program.

Conference Session

Organize Reporting Point-Solutions into a Data Warehouse

Casey Welch
Engineering Data Architect, Cerner Corporation
Drawing from the real-life experience of framing a "quick and dirty data mart" with a structure that can grow into an enterprise reporting and decision support solution, Casey introduced appropriate methodology and architecture. He defined where and how to start such a project AND how to incorporate the effort and intellectual capital hidden in existing reports. Focusing on a subject-area definition process that logically groups information delivery requests, the approach helps avoid "scope creep" that often dooms such projects.

Conference Session

Active Meta Data as an Enabler of the ODS within an Enterprise Data Architecture.

Leonard Dawson
IS Senior Analyst, &
Jerry McConnell
IS Senior Analyst, New York State Electric & Gas Corp.
Leonard and Jerry suggests that a single meta data repository can serve as the foundation for an Operational Data Store (ODS). Active, passive, and administrative meta data can be integrated into a single repository that includes business definitions, fields descriptions, and ODS data

acquisition process control rules and metrics. The ODS data acquisition process should be designed as meta data driven infrastructure on which new data can be deployed without the need for additional program code. Definitions and business rules, which are inherently volatile, should be designed into the active meta data so that they are isolated from the ODS data acquisition processing. A properly designed ODS data acquisition process can be used for daily updates, as well as for initial loads and conversions, restores, corrections, and data quality sampling.

KEYNOTE

The Future of Data Management

Bill Inmon
Partner, billinmon.com LLC
In a relatively short amount of time, the IT profession has come from punched cards and paper tape to petabyte data warehouses. And as technology has progressed, so has the stewardship and management of the technological environment. This presentation outlined how data management has evolved through the many different advances in technology over the years.

PHOTOS

Links to Photos

Event Photos can be found at:
http://www.wilshireconferences.com/MD2002/photos.htm

The DAMA/Meta-data Musical Jam at WINGS!!!!
http://www.wilshireconferences.com/MD2002/music.htm
These photos are from Stan Slossberg, who led a "band" of attendees to the open-mic night at a local music club called "Wings". The pictures are worth a thousand words – a good time was had by all.

Personal Tales:

1) A First-Timer

First-Person Experience of the DAMA International Symposium + Wilshire Meta-Data Conference

by Elizabeth Gray, conference attendee

This was my first DAMA Meta-Data Conference, and judging from my own experiences and the comments of others, it was successful on several fronts. I came away impressed with the range of offerings, all with an emphasis on tactical, practical applications.

The most common theme I noted was the evident maturing of the Data Management profession in general. The emphasis has definitely turned away from flashy tools and silver bullet solutions to the recognition that improvement can come only through careful, exacting analysis accompanied by sustained political support. The pace now is evolutionary, no longer revolutionary (if it ever was!). All organizations are working to build on now-familiar approaches—such as the Zachman framework—effectively in a corporate setting. The change is slow and incremental, and conference participants seemed to support numerous speakers' assertions: that is just the way it has to be, especially given the tremendous complexity of enterprise data management issues.

There was an interesting side note to all the discussion about measured change. Several presenters gave frank assessment of the errors they had made, and lessons learned over past years of investing in data management. In fact, for at least one speaker, several projects in a row were failures. Nevertheless, he articulated clear value from the work papers those projects produced, and they are providing a solid foundation for further efforts. The road is not as straight as planned, but the progress is still measurable. It was fascinating to hear these sincere, honest assessments.

I heard a lot about matching the approach to the particular organization. Many speakers repeatedly emphasized that what works in one setting may not in another. There is no "right way" to implement effective data management practices; each environment is unique and will require its own tailored approach. In fact, because these are often multi-year efforts, several project teams can expect flexibility in requirements and resources, based on the changing environment.

Last but not least, I perceived a lot of general satisfaction with the conference site, organization, and networking opportunities. Within the conference, everything went very smoothly. Registrations, breaks, AV equipment, room logistics, speaker logistics—as a participant I didn't notice a single problem. The atmosphere was authentically friendly, not off-putting. I saw lots of networking, particularly with the international participants. And I heard many people happily telling stories of what they done in and around San Antonio.

NOTE: My organization is small and just starting down the data management road. Consequently, I focused on core DM topics, and did not attend the many sessions offering information on the "latest and greatest" in DM, such as XML, new tools, or future trends.

2) Street Talk

Personal and collected observations from the recent conference experience.

By Paula Thornton

The Venue

The city was fabulous. Because of the locations of events and accommodations, most of our 'doings' centered around the San Antonio Riverwalk. With conference activities starting on Sunday, April 28th – the last (10th) day of the annual Fiesta San Antonio event – some attendees arrived early to catch a full day of activities on Saturday. They reported that the investment provided fabulous returns.

The central bulletin board at the conference was used not only as a 'feedback' mechanism of experiences at various restaurants (good/bad, price, format), but was also popular for coordinating social gatherings for chapter groups. The Chicago chapter coordinated one such dinner. Given seating accommodations on the long, narrow open balcony of one trendy restaurant, so many people 'crashed' the event that seats were added and people eventually were forced to join the group through the windows.

Perhaps unfathomable to northern residents facing inclement weather during this time-frame, weather in San Antonio was consistently in the high 80's and low 90's the entire time. Attendees took liberal advantage of breaks between sessions by going outside, that is, when they could break away from intense conversations in the hall. As the conference venue crosses into the 'central area' again, San Antonio is definitely worth repeating.

The Agenda

Sunday featured half-day Workshops, while Monday offered full-day tutorials. Both days were capped off with night school sessions. Given that I was trying to get a 'broad view' of the conference, I was glad that I gave myself 'permission' to hop through sessions. I made some rather unexpected discoveries.

Sunday

Chatting briefly with some of the attendees and in conversations with Bob Seiner himself, interesting feedback was gathered. While his workshop was titled, "Aligning Data Management Skills with Business Goals" many people attended his session specifically because of two references in his course description: knowledge/content management. Many data management people have recently found themselves thrust into roles where they suddenly have responsibility for these areas and needed help.

I 'had' to step into Zachman's presentation to see what he was up to (my last encounter having to present 'opposite' John, several years ago). While my first impression was MOS (more of same…not that there isn't always new blood that needs to hear John's core message unadulterated and from the source), I later went back and discovered some new items of note. While John consistently insists that defining an enterprise architecture is 'good business' he also reported that for government agencies it is now a matter of regulation. To gain budget approval, all major projects must clearly show how they fit into or expand the architecture (defining/creating architecture where there currently is none). Zachman's approach has been 'officially' adopted as the model for designing architectures. During a personal chat with John, he lamented that he was both frustrated yet stumped by adequate resolution to those who 'train' others in his approach, but miss the mark. Word of wisdom would suggest that learning from the master continues to be the best approach.

While titles that might have seem more "technical' didn't appeal to me, I was pleasantly surprised to stumble into Sridhar Iyengar's presentation on OMG models. The presentation was of less

value than simply having Sridhar impart his vast first-hand knowledge of the evolution and planned evolution of OMG and its related models. Several of us cornered him after his presentation for more details. I've actually been 'out of the loop' for a while (living in the Internet space) and was generally unaware of OMG's recent activities. Of specific significance was learning the relationships between various standards that we hear bandied about on a daily basis. While some argue for one approach over the other, Sridhar pointed out that some approaches clearly have no 'overlapping responsibilities' and therefore cannot take the place of one over another. It was evident that the future of OMG evolution is the focus on MOF (Meta Object Facility). The MOF modeling approach will eventually obviate the need for several current approaches (some may continue to engage in creating such models, but they will be largely redundant). Among those, surprisingly enough, would be UML itself. In the conversation after the presentation we hit on a significant conceptual context that Sridhar said he'd probably add to future presentations. Where XML is intended to create portable data and JAVA enable portable code, MOF will provide portable models/semantics.

Monday

The 'intensity' of the full-day courses was such that it made it difficult to 'surf' the offerings. I can say that the 'Graeme and Graham' show was truly entertaining. I can say this without suggesting that I agreed with everything they proposed, or the way they proposed it. I recommend we collectively dub them "Graham Crackers", an American tribute to 2 funny Aussies.

The line-up for the Night School sessions that evening was so enticing that I had 'highlighted' every session.

Tuesday

Starting with a keynote from Ed Yourdon, the main portion of the conference proved to be engaging. Standing out from the day of 'general interest' presentations was an engaging exchange initiated by Fabian Pascal. I highly recommend seeing for yourself what is being said by this man who claims that short of one recent product, there are no database products on the market that are truly relational. Failure to deliver the intended relational technology causes people to place blame on relational databases (such as performance constraints) that apply strictly to the technology design…not the concept. If appropriately implemented, the constraints largely experienced by most of us would not exist.

Wednesday

A key event of the conference was the morning's early focus on Achievement and Best Practices Awards. RBC Financial Group received the inaugural Wilshire Award for Meta-Data Best Practices, accepted by Moe Rifaie. Then, well-deserved recognition was given to Bill Inmon for his general contributions to the data management industry via Data Warehousing for many years and Davida Berger for her relentless and tireless contributions to planning conferences each year. Most stirring, however, was a lifetime achievement award presented to both John Zachman and his wife. Truly a 'speechless' moment (a rare phenomenon by his own account), John was emotionally touched by the moment and received a standing ovation from the audience, adding their concurrence to the recognition.

My Agenda

The rest of my comments are being moved into this section as they represent a truly 'slanted' view of the conference. I branched out of data management several years ago, and have made new discoveries about overlaps and virtually untapped synergies in other disciplines (see related comments at http://www.tdan.com/pubpers0110.htm). Some conference activities highlighted this focus.

On Wednesday, introduced by conference chairman Tony Shaw, one presentation was noted as being somewhat of a departure (read: experimental) from the typical topics. It was a sign to me that the convergence of like disciplines may begin to occur. Lynn Wojcik, Director of Content

Classification at Northern Light Technology gave a well-positioned introduction to the inclusion of Information Science (read: library science) techniques to meta data optimization. At the same time, Sandra Hostetter was in another room suggesting that until librarians and data administrators join camps and exchange ideas and approaches, data strategies will continue to miss their potential. Having been involved with practitioners from ASIS&T (American Society for Information Science and Technology) for the past 3 years, I have to concur.

Immediately following, there was an industry panel suggesting what they felt was the "Next Big Thing" for data management. Robert Seiner suggested that it would be knowledge/content management. Unscheduled addition to the panel, Len Silverston, got a rousing lone round of applause from me when he suggested that the next big thing was a need to focus on people issues. We both later commented that it struck us unusual that the comment received no response from the audience (we'll be studying this further to assess industry readiness for this concept).

Product and Tool Exhibitors

By John Baughman

The following is a snapshot of the companies which exhibited products and tools at the conference. Company names are hyperlinked to take you to their web sites.

ASG is a privately held global enterprise software and professional services firm with a full range of applications management, operations management, and information management services. Metadata aficionados will particularly appreciate the repository products.

Computer Associates (CA) offers data warehousing, knowledge discovery, and metadata management technologies combined with information visualization, infrastructure management technologies and CA Services. The Advantage Repository (formerly Platinum) is the market-leading repository product in terms of market share, and provides a comprehensive road map of the enterprise data warehouse and eBusiness information assets. CA is also the purveyor of Erwin, a staple to all data modelers.

Evoke Software helps companies develop a thorough understanding of their corporate systems prior to undertaking critical eBusiness initiatives. The Evoke Axio product suite provides a thorough, automated analysis of the content, structure and quality of data, and accelerates and improves the quality of eBusiness, CRM and data warehousing initiatives. Axio is a good tool for the serious analysis of the quality of legacy data.

Informatica provides enterprise analytic software that enables decision makers to transform business insight into competitive advantage. Informatica offers the industry's only integrated enterprise analytics suite, including a powerful data integration platform, cross value chain analytic applications, and real-time delivery of analytics via Web, wireless and voice. Informatica's mapping and visual assists make it extremely productive.

Knightsbridge is a systems integrator delivering high-performance "big-data[SM]" solutions. Their data warehousing and data infrastructure solutions empower organizations with either high data volumes or complex data challenges to leverage organizational data for competitive advantage.

Visual Studio .NET is Microsoft's rapid application development tool for building XML Web services and applications. Visual Studio .NET allows developers to rapidly design broad-reach Web applications accessible from any device and any platform. In addition, Visual Studio .NET is fully integrated with the .NET Framework, providing support for multiple programming languages and automatically handling many common programming tasks.

Data Junction Corporation markets data integration and data transformation technology. Data Junction provides transformation tools, developing solutions and services necessary to solve data and metadata integration problems. Data Junction assists in modeling processes, maps, etc while the engine provides a true jump-start to the integration process.

Experian helps its customers target customers, manage existing customer relationships and identify opportunities for profitable growth. Experian's Data Bases help customers evaluate and target customers and markets.

InConcept uses a variety of tools and techniques to help you design database systems according to your business requirements. They assist in data modeling and database systems, from analysis to implementation. They are distinguished by offering the *Journal of Conceptual Modeling* and training in various techniques like ORM.

PeerDirect products support the replication of relational information on most combinations of device and database, creating a single distributed infrastructure, supporting both relational

database and file system synchronization. PeerDirect products also let you manage the distribution of schema changes throughout a live replication network.

Trillium Software provides investigation, reengineering, cleansing and customer relationship matching functionality to a wide range of global organizations. Trillium provides companies a coherent way to standardize data and to provide a consistent, unified view of customers, products and services.

Ab Initio software is a platform for data warehousing, data movement, data transformation, and real-time analytics. Ab Initio supports application integration and parallel and distributed execution across heterogeneous networks, providing capabilities for ETL and EAI tasks within a single, consistent framework. This is a real step up from the hand-coded hodge-podge that many of us are trying to control.

ACME Software's DataLever is a visual data-programming environment that slashes the time required for data-centric development. It is a visual programming tool that makes it easy to create and run high-performance data-transformation engines.

Avellino provides comprehensive data integration and data quality software to help organizations leverage their mission-critical enterprise data for CRM, Data warehouse, ERP, & eBusiness initiatives. Avellino's Discovery suite automates data analysis and discovers data related problems by analyzing the content, structure, dependencies, relationships and quality of source data. Given the condition of most legacy data and associated metadata, any automated analysis tool is most desirable.

ChoiceMaker Technologies combines artificial intelligence technologies with software architecture to produce record matching/linking software. The Java architecture allows them to operate on any kind of data on any major database and to integrate with any kind of system. Their "Maximum Entropy De-Duper" is worth a good look.

Data Advantage Group, Inc.'s, MetaCenter™ is an analytical data management platform enabling companies to capture, analyze and manage metadata how and when they need to. It allows integration of metadata from all the disparate sources that most shops have and then makes it available via a browser.

DataFlux, a SAS company, provides technology for data cleansing, data augmentation, data consolidation and conversion. The tools assist in data transformation and merging to assimilate data from disparate sources.

DataMentors is a full-service data quality solutions company, providing a comprehensive suite of completely customizable data validation, transformation and database building products. The company offers a fully modular relationship matching and linking system that cleanses, organizes, standardizes and households data.

Embarcadero Technologies, Inc. has powerful database support and development solutions. Led by the success of DBArtisan, the database administration environment, Embarcadero now carries the same usable approaches to database design with ER/Studio and SQL development with RapidSQL.

Firstlogic's information quality software implements into CRM, ERP, BI, and data warehousing applications and helps businesses create a single record view by enhancing and matching customer (names and addresses) and business information (account numbers, SKUs, etc.). The data quality tools provide intelligent assistance for that difficult task.

Innovative Systems, Inc. helps organizations clean, understand and use data by providing data profiling, data quality and linking software and services. Our world-class solutions identify true

customers and detect non-intuitive relationships between all types of data. The result is a unified view of any customer or data element across the organization. Quality products will parse and standardize customer data.

Intelligent Medical Objects (IMO) leverages its extensive experience in electronic medical record and database technologies to provide vocabulary components, tool kits, and warehousing/data mining platforms for use within and outside healthcare. IMO's Adaptive Data Manager(tm) is one of the most powerful metadata modeling/database development environments built for Oracle users.

Melissa Data provides data quality solutions with emphasis on address and phone verification and correction, postal encoding, and data enhancements. Includes zip code and mailing list solutions.

Orenburg's BOARD Management Intelligence Toolkit (M.I.T.) combines the capabilities of traditional BI tools in a programming-free environment to deliver a framework for rapidly developed, customized and cost-effective analytics and decision support applications. The user can be as sophisticated as desired with this tool.

Rational Software provides a software development platform that improves the speed, quality, and predictability of software projects. This integrated, full life-cycle solution combines software engineering best practices, market-leading tools, and professional services.

Search Software America (SSA) develops and markets international software products that significantly enhance an organization's ability to search, find, match and group identity data, including names and addresses.