



DAMA MICHIGAN Bits & Bytes



Michigan Chapter of DAMA International

Summer 2004

Data: The Missing Link To Business Success **To Register: www.dama-michigan.org**

What is a DAMA Day?

It's a mini-conference planned by the local chapter where International Experts will discuss Data Management subjects like:

- Data Warehousing
- Data Quality
- Enterprise Data Planning and Data Management
- Business Intelligence including Customer Relationship Management
- Data Modeling/Architecture of Solutions

It includes vendor booths and attendees an opportunity to network with peers.

Program Speakers:

Larry English, international consultant on Information Quality (IQ) management and Information stewardship. Larry lives in Brentwood, Tennessee

Michael Brackett, international consultant on data warehousing, meta-data management and managing disparate data. Michael was also President of DAMA-I in 2000 and 2002. Michael lives in Olympia, Washington

Dr. Michael Cavaretta, an experienced practitioner and research scientist who is in deeply involved in the Data Mining practice that supports Ford Motor Company.

Kas Kasravi, an EDS practitioner, discusses the issues and strategies of getting control of information content trapped in unstructured data generally found in textual documents.

Michael Scofield, international consultant in data modeling, data warehousing and data architecture including data integration methods. Michael lives in Loma Linda, California

The Presidents Corner By Sid Stoffer

Experiencing the CDMP Exam

A couple of weeks ago, I and a number of fellow data fanatics took a beta (early-release) version of the CDMP certification exam. The Certified Data Management Profes-



sional (CDMP) exam is one of the specialty exams that can apply toward the Institute of Certified Computing Professionals (ICCP) certification. In addition to passing this exam with a score of 70%, the applicant needs to pass a "core" exam and another ICCP specialty exam with similar passing scores.

(Continued on page 2)

We hope by now we have attracted your attention. In the following pages, we will be providing details about the conference, location, speaker's bios, and more.

We hope you will attend. It is a wonderful change to listen to experts in our industry who are making significant contributions. See you there!

Editor

Corporate Sponsors



(Continued from page 1)

This was a risk-free test since it didn't cost anything other than shipping. However, the gory details were that I didn't really have a way to study for it. We were given a high level outline of topics covered a couple of days before the exam. It didn't occur to me at the time that I had in my possession an excellent reference. This reference is called ***Guidelines to Implementing Data Resource Management*** and is a series of data management related articles written by various well-known authorities in this field who also happen to be DAMA members. The book was compiled and edited by the Chicago DAMA chapter to be used as a fundraiser device for DAMA and DAMA Chapters. The bottom line is that I went into the test "cold" and scored a 69%. This was indeed frustrating (missing by 1%) and yet I felt good considering the wide range of topics covered and amount of highly cognitive thinking I had to do in the time allotted.

My company is encouraging the IT staff to get certified in any area that might be favorable to getting work. To this end, I have pursued and received several technical certifications in the last year or so. This has involved studying the material, taking some courses, buying books, and taking lots of sample exams in addition to the "real" certification exams.

Some questions need to be considered: "What is the business value of these kinds of activities?" and from a personal perspective, "Do these tests really provide additional opportunities for those willing to endure the pain of study and performance?" On the plus side, they do provide a point-in-time metric for an area of knowledge. From a company perspective, it could provide some evidence that a person can be considered someone worth listening to. From an individual perspective (at least in my case), I came out of the test with a better feel for areas that I am solid in versus needing more study. I definitely need

more study in data quality management. Some of the terminology and concepts are very specific and leave little room for value judgment. On the minus side, a company must have an interest or see value in a particular technical area for the certification to have value. For instance, some companies aren't currently listening to data management issues relative to other IT issues.

Additionally, my experience on some of the certification tests involves observing that some tests focus on minutiae of an area rather than mainstream concepts and practices. However, in the end, there are probably more pluses than minuses. Thus, I will continue plugging away at these kinds of activities. I would be interested in your thoughts on certification. Write us and we will include a follow-up article later in the year.

As a parting note, our first DAMA Day is on the horizon. We have an excellent program planned at the MSU Student Union, East Lansing, Michigan on August 10. We hope you can attend.

Sid Stoffer
President

Speaker's Biography

Michael Brackett

Mr. Brackett is the founder of Data Resource Design and Remodeling and was the President of Data Management Association (DAMA) International for 2000 - 2002. He helps public and private sector organizations develop an organization - wide integrated data resource, stop the continued production of disparate data, and resolve existing data disparity. He has written numerous books and articles and is a prominent speaker at national and international conferences. His recent book ***Data Resource Quality: Turning Bad Habits into Good Practices*** describes ten steps to stop the creation of disparate data. He is also a contributing au-

thor to the ***DAMA: Guidelines to Implementing Data Resource Management, Fourth Edition***

ISBN: 0-9676674-1-0.

Further information about Mr. Brackett can be found at members.aol.com/mhbrackett. Mr. Brackett can be reached at mhbrackett@aol.com.

Dr. Michael Cavaretta

Dr. Cavaretta is an experienced practitioner and research scientist who is in deeply involved in the Data Mining practice that supports Ford Motor Company's strategy in marketing, manufacturing, quality, and warranty subject areas.

Larry P. English

Mr. English, internationally recognized authority in information management and information quality, has consulted in 28 countries on five continents. He was featured as one of the "21 voices for the 21st Century" in the January 2000 issue of ***Quality Progress***. DAMA awarded him the 1998 "Individual Achievement Award" for his contributions to the field of information resource management. Mr. English's widely acclaimed book, ***Improving Data Warehouse and Business Information Quality***, has been translated into Japanese by the first information services organization to win the Deming Prize for Quality.

He is also a contributing author to the ***DAMA: Guidelines to Implementing Data Resource Management***.

Mr. English can be reached at the following address:

Information Impact International, Inc.
Phone: (615) 837-1211
Larry.English@infoimpact.com
<http://www.infoimpact.com>

Kas Kasravi

He is a valued member of the Elec-

(Continued on page 3)

tronic Data Services (EDS) team in Detroit, Michigan where he is involved in issues and strategies for controlling information content trapped in unstructured data. His area of expertise is text mining. Unstructured text is a growing area of concern for many companies as they attempt to comply with the Sarbanes-Oxley (SOX) requirements for document classification and retention particularly in the form of email. Also, it is very useful in mining information from web pages.

Michael Scofield

He is Assistant Professor, Health Information Management, at Loma Linda

2004 Board of Directors

President, Sid Stoffer,

EDS,
Phone: (248) 696-3118,
email: sid.stoffer@eds.com

Admin VP, Vernon Hoffner, Ph.D.,

Lawrence Technological University,
Phone: (248) 204-3058,
email: hoffner@ltu.edu

Secretary, Anna Whisenhunt,

EDS,
Phone: (248) 265-0004,
email: anna.whisenhunt@eds.com

VP of Programs, Nanette Yuzon,

EDS,
Phone: (248) 265-6426,
email: nannette.yuzon@eds.com

Treasurer, John Mieczkowski,

EDS,
Phone: (586) 575-4588,
email: john.mieczkowski@eds.com

VP of Membership, Jim Tepin,

State of Michigan,
email: tepinj@michigan.gov

VP of Marketing, W. Thomas Hamlin,

Extensible Information Systems (EIS)
Phone: (517) 887-3087,
email: thomas@thomahamlin.com

VP of Logistics, Teresa Sun,

Independent Consultant,
email: teresatsun@yahoo.com

VP of On-Line Services, Jerry Stembridge,

EDS,
Phone: (248) 265-6203,
email: jerry.stembridge@eds.com

University. He is a popular lecturer on topics of data quality assessment and data warehouse design. He has spoken at conferences and professional meetings in over 56 cities in the U.S. and overseas. His articles on data management have been published in various journals including *DM Review*, *InformationWeek*, *The Database Newsletter*, the *IBI Systems Journal*, and others.

Editor's Note: These gentlemen offer you insight and expertise in the leading issues in data management today. Normally, they are only available in large venues. DAMA-Michigan brings them to you at a very reduced cost at a local level. There is no reason not to attend.

Membership Report

By Jim Tepin

Vice President of Membership



James Tepin Jr.

To the prospective member,

For more information about becoming a corporate sponsor, corporate member, or individual member, please refer to our membership web page via our chapter site at: <http://www.dama-michigan.org>.

As for you students out there, please take note that our bylaws have changed. If your college or university is a corporate sponsor, you are entitled to the lowest rate charged for all DAMA events (if there is a rate charged at all).

Sincerely,

James Tepin
Vice President of Membership

Presentations By:

Michael Brackett

Overview of DAMA – International and relationship to local chapters

A brief description of the organizational structure of DAMA-I and how DAMA-I conducts its business. In addition, Michael will also explain how DAMA-I is changing its business model to be more proactive and responsive to its local chapters.

Overview of his book called “Turning Bad Habits into Good Practices”

The state of the data resource in most public and private sector organizations is low quality. The data are disparate and that disparity is rapidly increasing. Organizations are staking their future on a critical resource that is low quality and unstable. The author has written several books explaining detailed concepts, principles, and techniques for preventing data disparity and building an integrated data resource. This current book takes a less technical orientation by explaining the ten major groups of bad habits that lead to a disparate data resource, the impacts of those bad habits, how they can be turned into good practices, the benefits of the good practices, and the best practices to implement for early success. The primary orientation is a wake-up call to organizations about the bad state of their data resource, and to provide a direction for re-

(Continued from page 3)

solving the situation and creating a high-quality data resource that supports an organization's business.

(Excerpt copied from Michael's website <http://members.aol.com/mhbrackett/>)

Overview of Data Warehousing concepts and terms

Mr. Brackett will present some general data warehousing concepts and terms in the context of past and current engagements.

Dr. Michael Cavaretta

Data Mining

Dr. Cavaretta will presents in-cites gleaned from his work on applications of Data Mining, including marketing, manufacturing, quality, warranty, and strategy at Ford.

Larry English

Improving Data Warehouse and Business Information Quality

This presentation details the essential ingredients of an effective information quality environment for the data warehouse. It includes how to assess information quality (IQ) at the data sources and in the warehouse.



Written and edited by Matt Roush, Technology Editor
WWJ News radio 950, Detroit
mnroush@cbs.com

great lakes IT report

The Daily Tech News Source For Michigan & Beyond

**Start your day with the most informative,
 easy to read News Report on
 IT happenings in Michigan.
 Direct to your desktop five days a week.**

Free subscriptions available for a limited time.

**Comments? News tips?
 Want a free subscription for a friend?
 Email mnroush@cbs.com.**

**Marketing or advertising queries contact
 Wendy Baca, wjbaca@cbs.com**

Mr. English will describe processes for correcting defective data and for preventing recurrence of data defects. In addition, Larry will present both the technical and management requirements for a sustainable information quality environment for data warehousing

Kas Kasravi

Text Mining

Mr. Kasravi will discuss the issues and strategies of getting control of information content trapped in unstructured data generally found in textual documents.

Michael Scofield

Data profiling and data inventory (and metadata capture techniques)

Michael will explore the basics of surveying the data asset, and evaluating stable databases and data flows for their architecture and meaning. He will discuss domain studies and other techniques for discovering data anomalies, and the use of formatted dumps for evaluating anomalies. The presentation includes practical examples of data

(Continued on page 5)

(Continued from page 4)

problems in production databases, discussion of relationships with data stewards and application managers to help them understand their anomalies. Metadata capture techniques such as reverse data engineering used to discover or confirm data definition, quality, completeness, scope and consistency.

Facilities

The Student Union

The Student Union (The Union) is in the north end of Michigan State's main campus. It is one of the older buildings around. It is an access friendly building. It certainly has steps, but it also has an elevator to all floors. This is where the DAMA Day will be held. The meeting is self contained. Check our website, www.dama-michigan.org, for the most current version of the menu and to register.

Parking

Parking is always a challenge if you don't follow the rules. You need to remember that Michigan State University is home to one of the largest Criminal Justice programs in the country. As such, their cadets need something to do and issuing parking tickets is certainly a hometown favorite revenue generator. No

amount of pleading will persuade them to 'tear up the ticket'. If you get one, expect to pay it.

But there is hope. I would certainly recommend parking across the street (M-43 or locally know as Saginaw) and down one block (South) at East Lansing's main parking structure. It is bright yellow and orange. Walking to the Union, you will get a chance to window shop, see the sights, and get some exercise. For those who may wish to temp fate, you can obtain a 'one day parking pass' when you exit off US-127 North or I-496 at the Trowbridge Exit. Proceeding East on Trowbridge, you will encounter a small stand-alone building with a 'drive thru'. The last time I purchased a 'one-day parking pass' the cost was around \$4.00. You must still park in designated areas.

The East Lansing Parking structure is a better bet.

Bit & Bytes is Looking for Contributing Writers

If you have something to say and a talent for saying it, please contact the editor of Bits & Bytes. In our continuing effort to provide our readers with informative, educational and entertaining material we would welcome your contribution to our publication. The work doesn't pay much but we'll put your name in 10 point bold at the beginning of your article.

Looking forward to hearing from you.
Ed

Check Out the DAMA Michigan Web Site Links to:

Events

Data Management

Feature Articles

Membership

Contact Us

Members only Section

DAMA International

Bits & Bytes Newsletter

<http://www.dama-michigan.org>

For Web Site Information
Contact:

Vice President -
Online Communications
Jerry E. Stembridge
EDS (248) 265-6203
jerry.stembridge@eds.com

By Vernon
Hoffner
Comments:
hoffner@ltu.edu



What Have You Done About Data Quality Lately?

A previous column defined data quality and some easy to quantify examples of the cost of poor data quality. Quality data supports the decision making requirements at the operational, tactical and strategic levels of the organization to the degree necessary to make good decisions. Anything less results in the misallocation of corporate resources, creates dissatisfied customers, can be a root cause of unhappy and unmotivated employees, with a consequence of lower company profits and missed opportunities, and eventually the potential demise of the organization.

Subsequent columns provided some examples of ways to determine and measure the degree of data quality, and suggested some dimensions along which data quality could be measured and reported. It was suggested that measuring source system data quality should be one of the first steps in a data warehousing project. The quality of the data, or its lack, would be a good indicator of the potential success of the data warehousing project. [It

was also suggested that this would be good information for the development project team members, to help them decide if, or when, to have their resumes ready for a quick exit from a failing project.]

The question posed now is what have you done about data quality on your last or current project? Was the project the development of a data warehouse or a data mart? Was it a business intelligence, data mining or other decision support activity? Was it the maintenance or enhancement of an existing system? Was it the implementation of an enterprise resource planning system or other off-the-shelf packaged system? Was it a review/audit of financial and related systems to support compliance with Sarbanes-Oxley? Do some of these projects require more or less of an emphasis on data quality than the others? How would you rank these types of projects on the importance of data quality? Or, could you rank all of the projects at the same level of data quality importance, or lack thereof?

Data quality is equally important for all of these projects. Data quality is important in the development and implementation of a data warehouse or data mart. The data quality of the source systems will have a direct impact on the amount of data cleansing that will have to be performed in the staging area of the data warehouse. The amount of effort required to extract data from the

source systems, and then to create the data cleansing and data transformation processes in order to supply the data warehouse with quality data is estimated to be between 50% and 70% of the project time. But if we provide the data warehouse with erroneous data, the end users will discover the data problems and eventually stop using the data warehouse. The result is a failed data warehouse.

Data quality is important for a business intelligence or a data mining project. The following process exemplifies the typical result of poor quality data. In many cases, managers have to make decisions based on feelings and limited experience, instead of accurate data that precisely describes a situation or opportunity. The results of bad data is illustrated in Fig. 1.

Bad data in => bad information out
=> bad decisions => wasted re-
sources, missed opportunities =>
lower profits => lessened probabili-
ty of corporate survival.

Figure 1. Results of bad data.

Data quality should be an important component of a system maintenance or enhancement project. If we are making changes to an existing system we will have to test the complete system to insure that the changes satisfy the requirements and that no bugs were introduced into the system during the change process. If we are expending the resources to review, change, test

and implement enhancements to an existing system, we could expand the project to review the system's data quality and improve the data capture components of the system. The benefits of quality for operational and down stream decision making would probably far exceed the marginal cost of improving the data capture components of the system. Future data warehousing, business intelligence and data mining projects would have access to quality data.

Data quality is important part of any off-the-shelf package implementation, regardless of the size of the project. Can you be absolutely sure that all of the data capture and processing components of the vendor's system insures quality data? All you have to do is ask them – Right! Whether a new system is developed internally, developed externally, or purchased off-the-shelf, we still need to insure that only quality data can enter the system and be stored in its database. Who knows what the future holds for the use of this new data. We would probably perform extensive testing for developed systems, but the same rigor needs to be applied to any vendor's package.

Quality data is necessary for compliance to the Sarbanes-Oxley Act. The act makes the CEO and CFO personally responsible for accurate financial reporting. Inaccurate financial reporting can result in prosecution of the CEO and CFO the ut-

come could be significant fines and other punishments. They are very concerned procedures, processes and data that support the corporate financial reporting. Currently, any projects that are required to insure compliance with Sarbanes-Oxley will receive executive support. However, almost all of the business transaction processing systems will have some impact on financial reporting, and need to be reviewed/audited for activities and data that do not place the executives at risk of noncompliance. This mean that all projects involved with business application systems should contain a data quality component.

What did you do on your last/current project relative to data quality? Did you do all that could have been done? What could you have done more of or could have done differently to improve the quality of the data resulting from the project?

Come to the 2004 DAMA Day on August 10, in East Lansing to learn more about data quality and other data resource management requirements. Check the website www.dama-michigan.org for full details and registration concerning this event.

Dr. Vernon Hoffner is Professor of Information Systems in the College of Management at Lawrence Technological University. He can be reached at hoffner@ltu.edu.

Useful Web Links

Data Warehouse Institute

www.dw-institute.com

DM Review

www.dmreview.com

Bill Inmon

www.inmoneif.com

Intelligent Enterprise

www.intelligententerprise.com

Ralph Kimball

www.ralphkimball.com

The Data Administration Newsletter

www.tdan.com