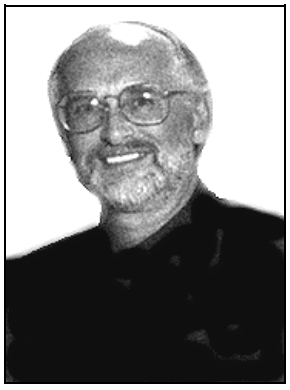
DAMA MICHIGAN Bits & Bytes

Michigan Chapter of DAMA International Fall 2002

Brett Champlin, Senior Lecturer Leads DAMA Michigan January Program Agenda

Brett Champlin is an internal Process Consultant with a large insurance company where he has been primarily responsible for building an enterprise process model solution including a Business Process Model Repository, process modeling methods, tools and techniques.



He has led over 50 projects redesigning business and IT processes there. Prior to that, Brett designed and implemented enterprise-wide architecture and software development projects for both very large and medium sized companies in a variety of industries.

Brett has over 20 years of experience working in Information Systems and Management. He has an undergraduate degree in Computer Science from Roosevelt University and an MBA from Illinois Institute of Technology (IIT). Brett is a Certified Systems Professional (CSP) and Certified Computing Professional (CCP) with proficiencies in Information Management, Systems Development, Software Engineering and Data Resource Management. He is a Senior Lecturer in the MBA and MSIS programs of the College of Business at Roosevelt University where he has taught for the past 16 years.

Brett is a popular speaker at international conferences, seminars and professional associations, including

(Continued on page 2)

Beyond The CMM:

Why Implementing The SEI's Capability Maturity Model Is Insufficient To Deliver Quality Information Systems In Real-World Corporate IT Organizations.

By **Brett Champlin**

The Software Engineering Institute of Carnegie Mellon University's Capability Maturity Model has become the de facto standard approach to managing information systems quality. Watts Humphrey's book sits on the desk of nearly every Software Quality Assurance professional throughout the county if not the world as their 'bible'. But, real-world applications have not produced the expected improvements. To implement even incremental improvements and move from one level of the CMM to the next takes years and seems to be difficult if not impossible to maintain. Why do we find ourselves in this situation? If the roadmap to

(Continued on page 2)

Jeff Wilkins, VP Strategic Solutions Practice, Netstar Corporation

When: November 12, 2002

Where: TRW Facility– Fowlerville

Presentation Title: Planning for Successful Data-Centered Initiatives

Presenter: Jeff Wilkins, VP Strategic Solutions Practice, Netstar Corporation

Running Time: 60 minutes.

Businesses recognize the potential value that can come from successful projects involving the acquisition and analysis of large volumes of data from disparate systems. For example, many companies are currently planning or implementing Customer Relationship Management (CRM) solutions with visions of reduced support costs, improved client retention and more effective marketing campaigns.

Unfortunately, industry literature is filled with horrifying statistics on the failure rate of IT projects. Executive management is interested in

(Continued on page 4)

Quality Information Systems In Real-World Corporate IT Organizations

(Continued from page 1)
quality, the CMM, is so clearly laid out, why do so many hard-working, intelligent professionals seem to be unable to achieve or sustain the quality in information systems that they desire?

The CMM was based on a simple model laid out by Philip Crosby during the quality movement of the late 80's. It was a generic process improvement model. The CMM translated that model to the software development process and added process specific detail to the model. It built a useful and valuable assessment tool. However, it is missing many of the components that go into delivering sustainable information quality in a corporate IT environment. This talk will discuss some of the many derivations and offshoots of the CMM and why they were mostly attempts to codify some of the missing pieces. One key missing is that the quality process improvement model assumed a degree of management oversight that doesn't generally exist in software development groups. Additionally, the other key critical components for delivering information quality are not even recognized in the CMM - data quality, systems capacity planning and integration, alignment of the IT portfolio with business strategy and needs, and more.

Not only is the CMM inadequate to ensure quality information systems, its proscribed timeframes for implementation are woefully out of step with the increased demand for speed in delivering new information functions in today's business environment. What can be done about it? How can we 'fix' this situation? The first step is to recognize that software development is not a stand-alone process and that we cannot hope to solve our real problems by isolating this process even further in our organizations.

Brett Champlin

(Continued from page 1)
the Institute for Operations Research and the Management Sciences (INFORMS), the Chicago Quality Assurance Association (CQAA), the Data Management Association (DAMA) International Symposia in the US and Europe, Meta Data Conferences, Data Quality Conferences, Delphi Group's Knowledge Management Conference, the Information Resource Management Association of Canada, and at DAMA chapters all over the US.

(Continued on page 8)

Fred Cummins, Enterprise Consultant, EDS

When: November 12, 2002

Where: TRW Facility– Fowlerville

Program: Evolution of UML

Presenter: Fred Cummins, Enterprise Consultant, EDS

.

Fred Cummins has been with EDS for 17 years. During that time he has worked on a variety of advanced development projects, including the development of a reflective object-oriented and logic-based programming language, an object-oriented application development methodology, a distributed computing framework and an engineering analysis management system involving workflow management and XML.

Since 1995 he has been the EDS representative to the Object Management Group where he has been involved in the development of a number of specifications including the UML Profile for Enterprise Distributed Object Computing (EDOC).

He has authored two books, most recently Enterprise Integration: An Architecture for Enterprise Application and Systems Integration. He is currently with the EDS Enterprise Information and Technology group focusing on web services technology, and he is co-chair of the OMG Business Enterprise Integration Domain Task Force.

November 12, 2002 DAMA Michigan Conference

Registration
12:00 - 12:30

Lunch/ Business Meeting
12:30 - 1:30

Jeff Wilkins
1:30 - 2:30

Break
2:30 - 2:45

Fred Cummins
2:45 - 4:00

Fall, 2002
By **W. Thomas Hamlin**
Comments:
thomas@thomashamlin.com

I am always amazed by the reporting requirements that operational systems cannot support. I guess that is the genesis of my interest in Data Warehousing (DW). I really don't like the term Data Warehouse or Warehousing. A warehouse is where product is stored prior to delivery to another process. I like the term Library much better. Libraries store relationships between objects (books and magazines by title, author and publisher) plus the object itself. By relating the scattered pieces together, a researcher can gain great insight into a problem. Why, because previously unidentified relationships between objects are discovered. New information is created and problems may be solved.

Very few people walk out their front door to stumble over Ed McMahon and the Publishers Clearinghouse Award team.

The enlightened reader may wonder if a discovery is by accident or design. The VAST majority of discoveries are incremental improvements leading to 'a discovery'. That is, many people are contributing to the knowledge base of an issue when the researcher discoveries a relationship that leads to better understanding. Very few people walk out their front door to stumble over Ed McMahon and the Publishers Clearinghouse Award team.

Let me put it another way using a little more humor. A drunk was leaning on a lighted lamppost on a very dark night. A foot patrol Officer stopped and asked what was going on. The drunk said he had lost his door keys and couldn't get into his house. The Officer asked if this is where he lost his keys. The drunk in amazement said no, he lost

(Continued on page 5)

DAMA Michigan Membership Report

By John Mieczkowski

Vice President of Membership

During 2001, we offered all new members, membership through 2002 for the same fee as one year membership. Most of our members got 3 to 9 months of additional free membership. Consequently, a majority of our membership is coming up for renewal December 2002 and notices are being sent as you read this. If you are a member under corporate membership, you won't get an individual renewal notice. However, we encourage you to make it known to your management, the benefits you have gotten from our meetings.

Since our first meeting with John Zachman in August, 2001, to our last meeting in September 2002 with Larry English, we have held eight meetings in all. One more meeting is scheduled November 12th of this year. Almost 250 have attended our meetings, averaging 30 people per meeting. All DAMA Michigan members also get DAMA-I membership included, and are entitled to discounts on their conferences and products also. Our membership drive for 2002 continues. Welcome to our newest corporate sponsor, Michigan State University of East Lansing, MI. MSU joined in May and upgraded their membership in July. Our newest corporate member is Pharmacia of Kalamazoo, MI with four new names for our roster, joining in October. We now have 168 members up by over 30 from January 2002. We have exceeded our goal for 2002, but much work remains with all the renewals coming up.

If you work for one of our corporate sponsors (EDS, GM, iITs, MSU or USI) then you are entitled to free membership under your company's corporate sponsorship. However, most of our corporate members also have room for one or two more

(Continued on page 4)

Jeff Wilkins to Address DAMA Conference Nov 12

(Continued from page 1)
obtaining the benefits of these large-scale, high visibility projects, but is gun-shy about moving forward with expensive IT endeavors.

This presentation explores the reasons why such projects fail and methods that can be used to overcome the obstacles that can lead to failure. It covers common issues that arise during systems integration projects where data is acquired from a variety of sources for analysis and action. Topics covered during the presentation include:

- Ø Obtaining and maintaining executive management commitment for the initiative.
- Ø Project execution strategies that lead to success.
  - o Organizational support.
  - o Project management.
- Ø Technical considerations that affect project success.
  - o Data quality.
  - o Legacy Applications: Migration vs. co-existence, standard technology vs. integration of varied platforms.
  - o Timely Updates: Real-time vs. batch.
  - o Connectivity: Dial-up vs. LAN/WAN, protocols, security and reliability.
  - o Operational readiness and capacity planning.
  - o Problem resolution, escalation, diagnostics and support.

For additional information, please contact:
Jeff Wilkins
VP Strategic Solutions Practice – Netstar Corporation
734-954-0648
jwilkins@netstar-corp.com

Coming November 12 "The Evolution of UML"

During his November 12 presentation Fred Cummins will discuss the evolution of UML and its importance to the industry. He participated in discussions when OMG undertook the development of the UML standard, and was involved in early work in OMG leading to the Model Driven Architecture (MDA) strategy. UML "profiles" have been and will continue to be developed for specialized applications of UML. In the long term, UML should support integrated modeling of systems with a variety of viewpoints supporting the work of specialists. The UML Profile for Enterprise Distributed Object Computing, the UML Profile for Enterprise Application Integration, and the Common Warehouse Metamodel are steps in that direction.

Membership Report

(Continued from page 3)
names, so contact me and I will see about adding you to our roster. If your company is not on our list then individual membership is only $45.00 and you get a vote. If you have one or more coworkers interested, then I would be glad to help make the case for corporate membership with your assistance.
I welcome your comments or questions regarding DAMA Michigan membership.

John A. Mieczkowski
VP of Membership, DAMA-Michigan

DAMA Michigan

Chapter Membership Information Contact

John Mieczkowski

V P Membership

(248) 874-4058

John.Mieczkowski@eds.com

DAMA MICHIGAN Bits & Bytes

8 Very Good Reasons To Join DAMA Michigan?

1. Professional Programs
2. Professional Networking
3. Education Programs
4. National and International Speakers
5. Resource Library
6. Professional Collaboration
7. Compare Notes with other DM Professionals
8. Enjoy a Pleasant Meal with Colleagues

Join Today!
Contact John Mieczkowski at 248-874-4058
or

www. dama-michigan.org

Point of View

(Continued from page 3)

his keys across the street. 'Why then', said the Officer, 'are you looking for them here?' 'Because', said the drunk, 'this is where the light is!' The moral to the story is that without background research to shed some light into the darkness most of our searches are in the wrong place.

This brings me to my topic for this issue. I am working on an Operational Data Store (ODS) project for a large retail merchandiser. The ODS will eventually feed their DW, a near-real time operational reporting system, and heaven knows what other future downstream systems. The data sources are the legacy operational system and the new ERP system.

Now catch your breath and read on. Can you imagine some of the issues associated with the project? For starters, the interface (Extract, Transform, Load (ETL)) team wants everything from legacy 'as is'. No effort is being made to update or improve the legacy data restrictions/limitations. There are no plans for incremental increases of understanding into the data. For example, their 'human being' table captures a person's name as follows. Refer to table 1.

Field Name	Field Type	Field Length
First_Name	Varchar	15
Middle_Intial	Varchar	1
Last_Name	Varchar	25

Today, whether we like it or not, we employ, work with, marry, educate and sell products and services to people from many nations and cultures. And guess what, they don't necessarily do things the way we do. Personal names are a good example. The experienced reader may say so what, we are lucky to get anything from legacy at all. Downstream applications can solve their particular problems when they have time. Good point, but a better point is that data designs, especially bad ones, quickly become enterprise wide standards. Accepting the data 'as is' is a prime example of this philosophy. What would it take to redesign the length of the fields to 50? What about additional fields to capture even more information and meet international standards? Moreover, this system would be unacceptable to me as a customer/employee since, my 'legal name' is William Thomas Hamlin but I prefer W. Thomas Hamlin. When I was younger (no comment from Bill Hepburn) there were too many 'Bills' in my family. My sainted Mother began to call me Thomas and it stuck.

Secondly, getting 'everything' with regards to data means the project

doesn't have firm business requirements. In the defense of the business types, we need to examine Hamlin's First Corollary, which states '*it is impossible to predict the future reporting requirements of an organization*'. Most projects simply duplicate existing reports and call it quits.

This is why DWs were created, that

(Continued on page 6)

How Dirty is Your Data

How Dirty is Your Data
Vernon Hoffner

New corporate financial reporting legislation requiring CEO's and CFO's to sign-off on the authenticity of the numbers contained in their corporate financials has heightened the stake the executives have in the accuracy and completeness of their organizational information. This should mean that corporate accounting data will be "cleaner" now, or at least the numbers that have to be reported to the public would be reasonably accurate. But what about "other" data that we use for decision making?

It has been said that data is a corporate asset. Since data is an organizational asset, we need to insure that we have quality data. We need to obtain management support of quality data management initiatives. We need quality data in our data warehouses and data marts in order to make good business decisions. In addition, we have read examples of horrendous data quality that have caused significant problems for data warehousing and business intelligence projects.

What can we do? Practice what we advocate! Make decisions and recommendations based on data. One of the first steps of a data warehouse project should be to analyze the data that will be used to determine its level of quality and to determine the clean data processes that should be implemented to obtain reasonably accurate data. This could also be considered the CYA step [:=))] of a data warehouse project, since the extracting and transforming of source system data into quality data that can be loaded into a data warehouse is usually the most time consuming phase of the project.

Analyzing character data can take several forms. The simplest is just a frequency of all of the different values for a particular field. This works fine when there is a limited number of values in the field, for

(Continued on page 7)

Point Of View

(Continued from page 5)

is, to offer a flexible time invariant data repository of the corporate 'truth' for 'a priori' and 'ad hoc' queries. An 'a priori' query means I have a very good idea what I am looking for because I have done the background work to understand a problem and need the final piece to solve the problem.

An 'ad hoc' query could be a poorly planned 'fishing expedition' where you figuratively throw a hook in the water to see what you might catch similar to what the drunk was doing in the humorous vignette above. DWs are a good place for incremental improvements.

Now for the final point in this issue, what is a good indicator that a DW project is successful? It is not what is the accepted norm for traditional operational system projects. Traditional projects do not look for immediate change requests or for more work to be done. For those projects immediate change requests are a signal that the original system requirements were not sufficiently understood. However, the best indicator of success for a DW is request for more! More reports, more data, more SQL, more of everything is needed. This is good because it means that the users understand the value in the DW and more importantly they are using the DW. Think of this as job security in a poor IT market.

DAMA MICHIGAN
Bits & Bytes

Submit articles for publication in Bits & Bytes to:
Bill R. Hepburn—Editor:
Newsletter_editor@dama-michigan.org

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

(Continued from page 6)

example, sex codes where we would be expecting only an "M" or "F", any other values, or missing values would be in error. An alternative for fields that have a larger number of values would be to compare the field to a table containing all of the permitted values, identifying those values that do not match the table entries. Any missing values would also be counted as errors. This provides a good first estimate of the percentage of errors in character fields that have a limited domain of permitted values. If the field has an unlimited set of potential values, such as, customer name, then we can use the fuzzy logic capability of one of the data quality software packages to identify potential duplicates.

The final step in determining how dirty our data is, would be to audit a sample of the records that contain correct data.

Numeric fields require a different type of analysis. We can test for missing values in those fields that should have a value. We can test the field against a permitted range when that is possible. Are zeros or negative quantities allowed in the field? An example would be the hourly rate for an employee in personnel data. The company would have specific minimum and maximum hourly rates for all employees and all rates should be within this range. If there is no maximum or minimum value for a field, then we can calculate the average and standard deviation for the field. Records with values beyond a specified standard deviation boundary, for example, four standard deviations, would be candidates for review and inclusion in the count of potential errors.

The above should identify the obvious single field errors. Next we can check for multiple field relationships that can be used to identify inconsistencies. An example from the health care field would be combining the diagnosis field and sex field to identify invalid combinations. A diagnosis of pregnancy for a man, or enlarged prostrate for a woman are errors for these two combined fields. Depending on the subject orientation of the data warehouse, a number of multiple field combinations may be available for validation. We just have to ask the "What IF" question of possible field combinations to identify the candidates for evaluation.

At this point, we will have identified the records that contain fields that have values that are not correct. The fields contain values that are not in the domain or range of permitted values for that field. The next step is to determine the accuracy of the data. In this case I am defining accuracy as free from mistakes or errors and correctness as data values for the field that are within the permitted domain. An example would be the situation where an employee's birth date is recorded as January 2, 1970. This is a correct as a possible or permitted value for a birth date. However, this would be an inaccurate birth date if the actual employee birth date were October 3, 1971.

The final step in determining how dirty our data is, would be to audit a sample of the records that contain correct data. Depending on the subject area of our proposed data warehouse, a relatively small sample of the clean records can be selected for the accuracy audit. Then we can verify the accuracy of each record by comparing the recorded values with the field values of the original source documents.

When we have completed this process, we will have determined how dirty our data is and we will have a data based foundation for estimating likelihood of success and/or the amount of effort required to ensure that the resulting data warehouse will contain quality data.

What do you think? Let me know at hoffner@ltu.edu.

Dr. Vernon Hoffner is Professor of Information Systems in the College of Management at Lawrence Technological University.

DAMA Michigan *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

Coming Events

Meeting Date	Meeting Time	Guest Speaker	Topic	Location
Nov. 12th	12:00 Noon—4:00 P.M.	**Jeff Wilkins** **Fred Cummins**	**Planning for Successful Data-Centered Initiatives** **Unified Modeling Language**	TRW 500 E. Van Riper Rd Fowlerville, MI
December		No Meeting.		
Jan. 14th	12:00 Noon—4:00 P.M.	**Brett Champlin**	**Beyond The CMM**	Siemens Dematic Grand Rapids, MI

Brett Champlin

(Continued from page 2)

He is a Past President of the Data Management Association (DAMA) Chicago Chapter and has been on the board of directors of DAMA International for the past 5 years.

Brett is a director of the Institute for the Certification of Computing Professionals (ICCP) and chairs their strategic planning committee. He was a contributor to the publication, "Guidelines for Implementing Data Resource Management", produced by DAMA Chicago and published by DAMA International. He is an occasional contributor to DM Review and TDAN (The Data Administration Newsletter).

Brett will top the agenda for DAMA Michigan's January 14, 2003 meeting in Grand Rapids. The meeting will be held at the Siemens Dematic facility from 12:00 Noon to 4:00 P.M. Lunch and refreshments will be provided.

After Thoughts