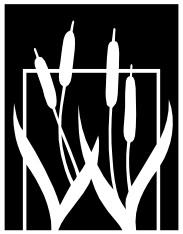


Whitemarsh
Information Systems Corporation

Data Management Capability Maturity Model

Whitemarsh Information Systems Corporation
2008 Althea Lane
Bowie, Maryland 20716
Tele: 301-249-1142
Email: Whitemarsh@wiscorp.com
Web: www.wiscorp.com

Table of Contents

1.0 What is a DM/CMM

Characteristic	Description
Definition:	A model of and then a process for determining the level of maturity that exists within an organization's data management environment.
Scope:	Scope ranges: • From an individual project to program. • From contained function to a community of interest. • From a project organization to an entire enterprise.
Objective	Determination of an empirical measure of data management maturity such that: • It can be tracked • Can be employed to construct specific prescriptions for data management maturity improvements within costs, time, and ROI constraints.

Characteristic	Description
Value in Accomplishment:	Enables organizations to improve their data management function so that they can increase productivity, increase quality, decrease cost, and decrease risk.
Minimum Characteristics:	The model of and the process for assessments must be valid and reliable. It must be malleable to meet local needs and special situations.

Fundamental Architecture for any Capability Maturity Model

(slide courtesy of Bret Champlin)

Process Maturity Level	<==========Categories==========>			
	Organization	Project	Process	Technology
	◆ Policy ◆ Resources ◆ Oversight ◆ Communication ◆ Training	◆ Planning ◆ Tracking ◆ Control ◆ Subcontracting	◆ Definition ◆ Execution ◆ Analysis ◆ Control	◆ Insertion
Level 1 Initial/Chaotic	Unclear relationship to work processes; unplanned commitments	Unpredictable cost, schedule, quality, and performance	None – undefined, no controls; no change management	Incidental
Level 2 Repeatable	Attention to planning and resource allocation; cost estimates; commitment management and change control	Highly variable costs; highly variable quality; reasonable control of schedules	Informal and ad-hoc methods and procedures; performance tracking	On project-by-project basis
Level 3 Defined	Attention to goal, process and measurement definition and communication; policies clearly defined related to process and results	Project management methods and standards defined; Reliable costs, schedules; Improving but unpredictable quality	Process methods and standards defined. Roles defined.	Project Management automation
Level 4 Managed	Attention to training, quality planning; quality policies and training	Project quality measurement (quantitative quality plans, goals, and measurements for results)	Quantitative statistical control (process quality measurements defined, planned and tracked).	Process Management automation
Level 5 Optimized	Attention to productivity planning, training	Refinement of project standards; error tracking and analysis	Continuous Process Improvement; productivity plans, standards, and tracking	Economically justified technology investment; Information technologies fully integrated into all phases of work.

What Is Being Assessed Are Your Key Processes

- Each Level Has Several Key Process Areas
- Each Key Process Area Has Goals
- Each Goal Has a Set of Activities.

Key Issues within this Presentation

- How do you know, determine, or cause a CMM to be valid. That is, that it measures what is needed to be measured?

- How do you know, determine, or cause a CMM to be repeatable. That is, its results are independent of assessor, program, project, or organization?

- How can you know what to do, when to do it, and what benefit you will receive if you act on computed level results?

2.0 What Are the Components of a DM/CMM Environment?

Component	Description
Generator	A data model and supporting computer-based information system for creating, evolving and maintaining assessment instruments.
Assessor	A data model and support computer-based information systems for conducting assessments and for creating prescriptions for future action based on assessments. A supporting function within the Assessor System.

3.0 Prior DM/CMMs

- Craig Mullins
- Bert Parker and the MITRE Model
- Bret Champlin's Research
- Data Blue Print DM/CMM (i.e., DM3)
- Others

3.1 Bret Champlin's Research

- Data Categories MM (BC Ministry of Forests)
- Data Management Maturity Model (Agosta)
- Data Management Maturity Model (Dravis)
- Data Management Maturity Model (MITRE)
- Data Management Maturity Measurement (Aiken)
- Data Resource Management MM (Champlin)
- Data Warehouse Information Management MM (Ladley)
- Data Warehousing MM (Marco)
- Der Reifegrad des Datamanagements (Schnider)
- Enterprise-Wide Data Management Maturity Model (Parker)
- Information Delivery Maturity Model (Computer Associates)
- Information Evolution Model (SAS)
- Information Maturity Framework (BC Ministry of Trans & Hwys)
- Information Quality Maturity Model (English)
- Meta Data Management Maturity Model (Stephens)
- Stages of an Active Data Warehouse (Brobst)
- Stages of Growth (Nolan)

Will the Real Silver Bullet Please Stand Up !

Oops, We're in Trouble...

3.2 Bert Parker and the MITRE Model (1994, The MITRE Corporation)

Bert Parker at The MITRE Corporation		
Level	**Name**	**Key Process Areas**
1	Initial	<absence of repeatable, defined, managed and optimizing DM processes>
2	Repeatable	Data Requirements Management Data Project Planning Data Project Tracking & Oversight Data Contract Management Data Quality Assurance Data Configuration Management
3	Defined	Organization Process Focus Organization Process Definition Training Program Integrated Data Management (Project Mgt) Data Product Engineering Inter-Group Coordination Peer Reviews

Bert Parker at The MITRE Corporation		
Level	**Name**	**Key Process Areas**
4	Managed	Quantitative Process Management (data gathering and analysis) Quality Management
5	Optimizing	Defect Prevention Technology Change Management Process Change Management

3.3 Craig Mullins (http://www.tdan.com/i003fe04.htm)

Craig Mullins (TDAN, 5/2000)		
Level	**Name**	**Description**
1	Initial	• No strict rules or procedures regarding data management. • Data may exist in multiple files and databases; using multiple formats (known and unknown); and stored redundantly across multiple systems (by different names and using different data types). • No apparent method to the madness and few, if any, attempts have been made to catalog what exists. • Changes are made "on the fly" as they are requested by program development. • If a centralized data management group exists, it functions merely to apply the change requests • The quality of data depends on the skills of the technical programmer analysts and coders. • Organizations will take on monumental tasks with little knowledge of their impact causing project cancellations

Craig Mullins (TDAN, 5/2000)		
Level	**Name**	**Description**
2	Repeatable	● Adhere to a data management policy. ● Policy dictates how and when data structures are created, changed, and managed. ● Rely on a central person or group to understand the issues and implement the data structures of the organization ● Database administration function.
3	Defined	● Documented and established a data management policy as a core component of their application development lifecycle. ● Policy is enforced and testing is done to ensure that data quality requirements are being met. ● Organizations typically understand the business meaning of data ● Created a data administration function to augment database administration. ● "Data is treated as a corporate asset"

Craig Mullins (TDAN, 5/2000)		
Level	**Name**	**Description**
4	Managed	● Managed metadata environment. ● DA and DBA groups catalog and maintain metadata for corporate data structures. ● Application development and end-user staffs access, use, and benefit from metadata ● Data management group involved (at some level) in all development efforts
5	Optimizing	● Continually improve the data access, data quality, and database performance through audits and process change ● No change is ever introduced into a production data store without it first being scrutinized by the data management organization and documented within the metadata repository. ● Data Management process are continuously assessed and improved.

3.4 Data Blue Print DM/CMM (i.e., DM3)

<table>
<tr><th colspan="3">Data Blue Print Corporation (M. David Allen's Masters Thesis, VCU (2004)</th></tr>
<tr><th>Level</th><th>Name</th><th>Key Process Areas Description</th></tr>
<tr><td>1</td><td>Initial</td><td rowspan="5">● Data Program Coordination
● Enterprise
● Data Integration
● Data Stewardship/Quality
● Data Development
● Data Support Operations

Note: Each KPA then has a set of questions. Each question has five characteristic answers that result in a Level-based assessment.</td></tr>
<tr><td>2</td><td>Repeatable</td></tr>
<tr><td>3</td><td>Defined</td></tr>
<tr><td>4</td><td>Managed</td></tr>
<tr><td>5</td><td>Optimizing</td></tr>
</table>

4.0 Clear and Present Need

- Should DM/CMM Be a Silver Bullet?
- How Can Validity, Reliability Be Achieved?
- How Can Extensibility Be Achieved

4.1 Should DM/CMM Be a Silver Bullet?

- Absolutely not
- If we cannot agree on definitive canonical data models, how could we possibly agree on a style, function, and organization-based process model?
- Technology affects what and when we do data management processes
- We must have the ability and "right" to evolve and maintain as needed
- It's "our" processes that are being assessed, not somebody else's

Besides, the Lone Ranger is dead.

4.2 How Can Validity, Reliability Be Achieved?

Validity

- There needs to be consensus on the key products of data management (what, not how)
- An Internet based identification, selection and ranking seems appropriate
- There does not need to be consensus on the key processes
- Industry consultants, DAMA, et al should participate in determination

Reliability

- There needs to be a way to "test" a DM/CMM across a wide swath of organizations about which the "maturity" is already known.
- There needs to be sufficient statistical analyses run to determine internal consistency
- An Internet based DM/CMM administration and scoring seems to be appropriate

4.3 Requirements for a DM/CMM

- The DM/CMM must be able to be generated just like a "Data Warehouse."
- Dimension Tables of Key products, et al
- "Fact" tables that are associations across dimensions
- Ability to stylize the actual questions
- Ability to administer the DM/CMM from project to enterprise
- Ability to track iterative administrations
- Ability to correlate against metrics
- Ability to set within idealized and "real" methodologies

And, We must CMM the DM/CMM

DM/CMM Characteristics		
Level	**Name**	**Description**
1	Initial	It is ad hoc. Everybody has their own. It's proprietary, not public.
2	Repeatable	It becomes independent of the person or organization administering.
3	Defined	It is fully documented, determined by subject matter experts and validated
4	Managed	Results are stored and cross-related within and between organizations. Predictions are able to be made, and they're correct.
5	Optimizing	As time, resources, technology et al changes the DM/CMM is able to be adjusted to fit new needs, environments, etc.

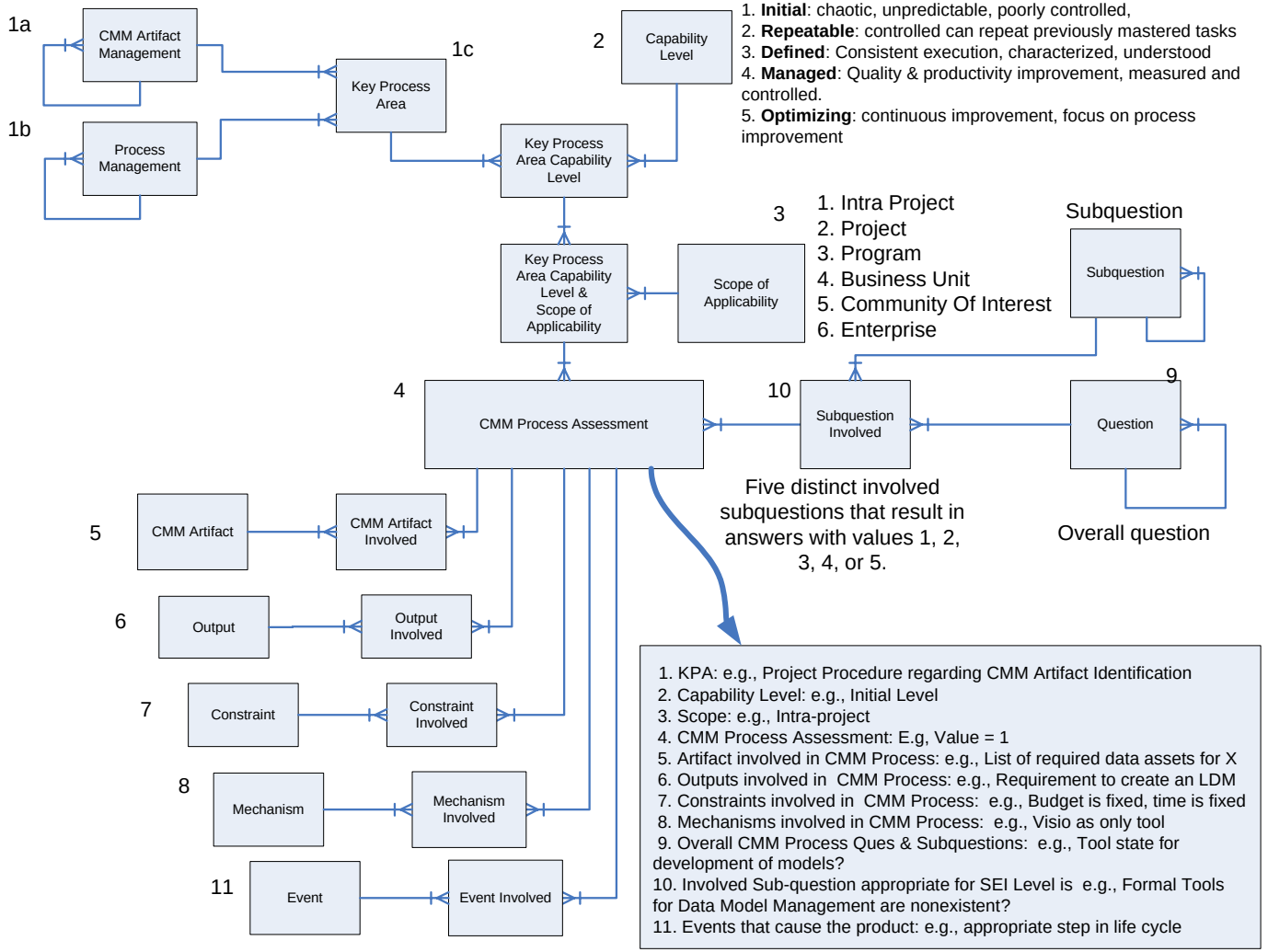
5.0 Engineering the Components of a DM/CMM

DM/CMM Component	Description
DM/CMM Data Models	Generator Data Model of artifacts that are involved in any DM/CMM assessment Assessor Data Model of artifacts that are captured and analyzed when an actual DM/CMM assessment is being conducted at an organization and/or on a project.
DM/CMM Process Models	A supporting information system and manual processes to assist in the formulation of the actual questions across the levels that will comprise the assessment instrument. A supporting information system and manual processes to assist in the capture, analysis, and reporting of any given assessment.

5.1 Generator System

- Generator Data Model Diagram
- Generator Data Model Tables
- Generator Data Model Processes

Generator Data Model Diagram



Capabilty Maturity Model Data Model
January 10, 2006

Generator Data Model Tables	
CMM Dimension Table	**Description**
CMM Artifact Management	Specific process associated with the management of an artifact that is being. Examples include: • Type Management (identification, creation, and management) • Instance Management (identification, creation, management, and disposition)
Process Management	Specific process associated with the management of the process associated artifact management. Examples include: Project Management (Identification, creation, and management) Program Management (Identification, creation, and management
Capability Level	The specific name associated with the assessed level. CMM levels are: initial, repeatable, defined, managed, and optimized.

Generator Data Model Tables	
CMM Dimension Table	**Description**
Scope of Applicability	Domain of the target of the assessment. Examples are Intra-Project, Project, Program, Business Unit, Community of Interest, and Enterprise
Question	Essential basis for the question. Examples include: Guidance, processes, automation support, history, Repeatability, Milestones, Strategies, Controls, Reversability, Configuration Management, and action plans
SubQuestion	Essential characterization of the assessed target of the questions. Examples include: organizational, scope, level of importance, accomplishment, and maintenance.
CMM Artifact	Identification of artifact that is involved in a process that is assessed. Examples include: [ISO 11179] data elements, database domains, database object classes, logical data models, physical data models, etc.

Generator Data Model Tables	
CMM Dimension Table	**Description**
Output	Identification of the product that is created, evolved or changed by an assessed process. Examples include: requirements, general design, detailed design, Tests, Documentation , etc.
Constraint	Identification of an item that affects the ability of an assessed process to accomplish its stated purpose. Examples include: budget, deadline, and methodology, deliverables.
Mechanism	Identification of an assessed process involved tool or facilitator. Examples include: Excel spread sheet, word processor, Visio-like tool, ER modeler, CASE tool, and metadata repository.
Event	Identification of an event that is involved with the assessed process. Examples include crisis response, life cycle milestone, and review.

Generator Data Model Tables	
CMM Association Table	**Description**
Key Process Areas	The intersection of CMM Artifact management and process management. A resultant association would be "Data Management Policy Management deals with Data Asset Instance Identification" or "Data Management Policy Management deals with Data Asset Type Identification."
Key Process Area Capability Level	The intersection of Key Process Area Capability and Capability Level. A resultant association would be "Data Management Policy Management deals with Data Asset Management is assessed to be Repeatable."
Key Process Area Capability Level Scope	The intersection of Key Process Area Capability Level and Scope of Applicability. A resultant association would be "Data Management Policy Management deals with Data Asset Management is assessed to be Initial has a scope of Enterprise."
Subquestion Involved	The intersection of question and a subquestion. A resultant association would be "Automation Support has the context of Level Of Importance,"

Generator Data Model Tables	
CMM Association Table	**Description**
CMM Process Assessment	The intersection of Key Process Area Capability Level Scope of Applicability, and an involved subquestion. A resultant association would be "Processes has the context of Level Of Importance has an applicability of Data Management Policy Management deals with Data Asset Type Identification."
CMM Artifact Involved	The intersection of a CMM Artifact and a CMM Process Assessment. An example would be: "A [ISO 11179] Data Element affects Guidance has the context of Organizational has an applicability of Data Management Policy Management deals with Data Asset Type Identification."
Output Involved	The intersection of a output and a CMM Process Assessment. "An output affects Guidance has the context of Organizational has an applicability of Data Management Policy Management deals with Data Asset Type Identification."

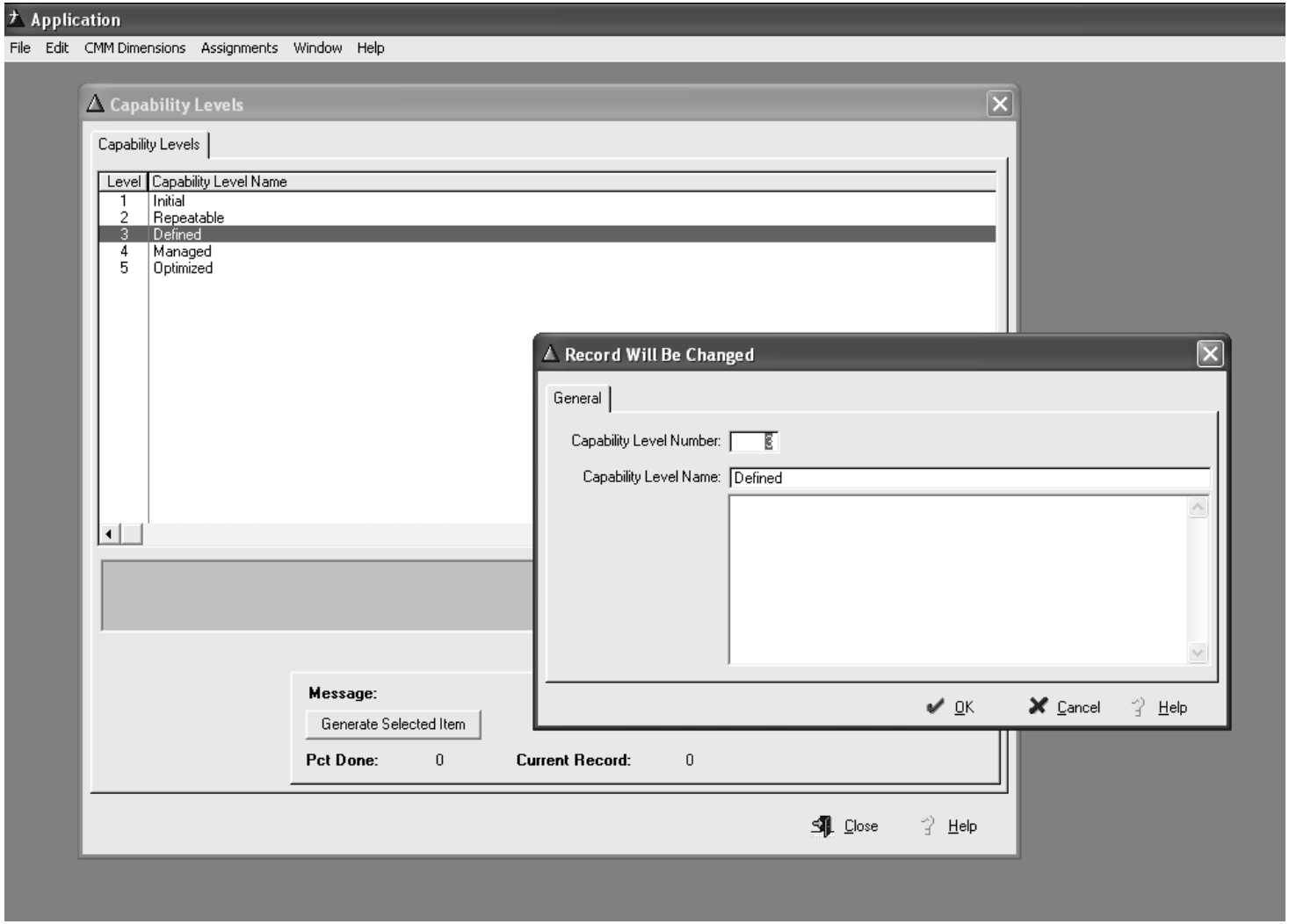
Generator Data Model Tables	
CMM Association Table	**Description**
Constraint Involved	The intersection of a constraint and a CMM Process Assessment. "A constraint affects Guidance has the context of Organizational has an applicability of Data Management Policy Management deals with Data Asset Type Identification."
Mechanism Involved	The intersection of a mechanism and a CMM Process Assessment. "A mechanism affects Guidance has the context of Organizational has an applicability of Data Management Policy Management deals with Data Asset Type Identification."
Event Involved	The intersection of a event and a CMM Process Assessment. "An event affects Guidance has the context of Organizational has an applicability of Data Management Policy Management deals with Data Asset Type Identification."

"Inherited Meaning" for an instance of the CMM Assessed Process Row		
ID #	**Inherited Associated Table**	**Example Meaning**
1	Key Process Area	Project Procedure regarding CMM Artifact Identification
2	Capability Level	Initial Level
3	Scope of Applicability	Intra-project
4	CCM Process Assessment	Value = 1
5	Artifact involved	A specific artifact,
6	Outputs involved	Requirement to create an LDM
7	Constraints involved	Budget is fixed, time is fixed
8	Mechanisms involved	Visio as only tool
9	Question & Subquestions:	What is the tool state for development of model
10	Involved Sub-question	Is the formal tool for data model management nonexistent?
11	Event involved	Data model review step

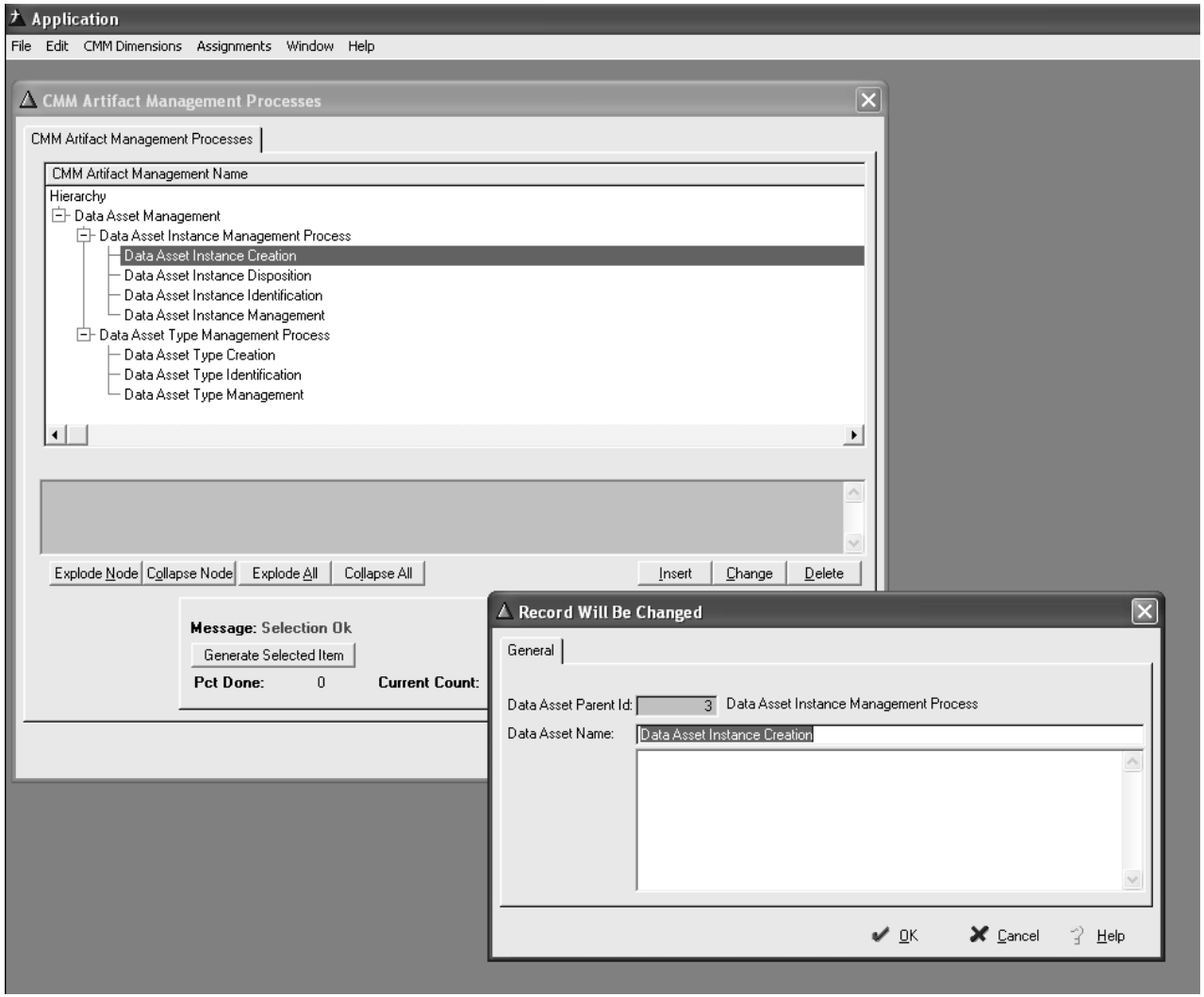
Generator Main Processes

- Dimension Processes
  - Flat Dimensions
  - Recursive Dimensions
- Assignment Processes
- ReAssignment Processes
- Deletion Processes
  - Dimensions
  - Associations

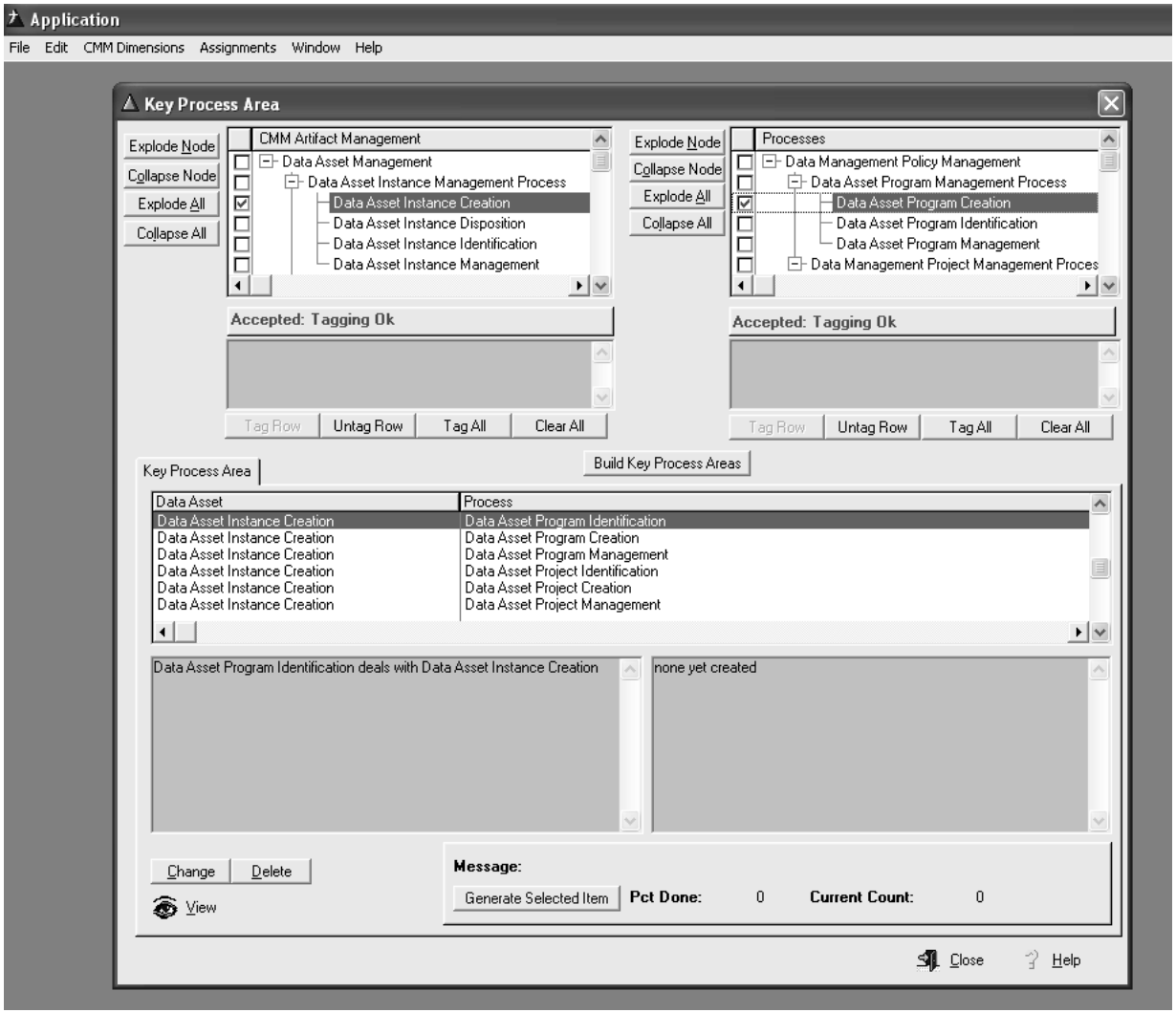
Flat Dimension Creation and Selected Association Generation



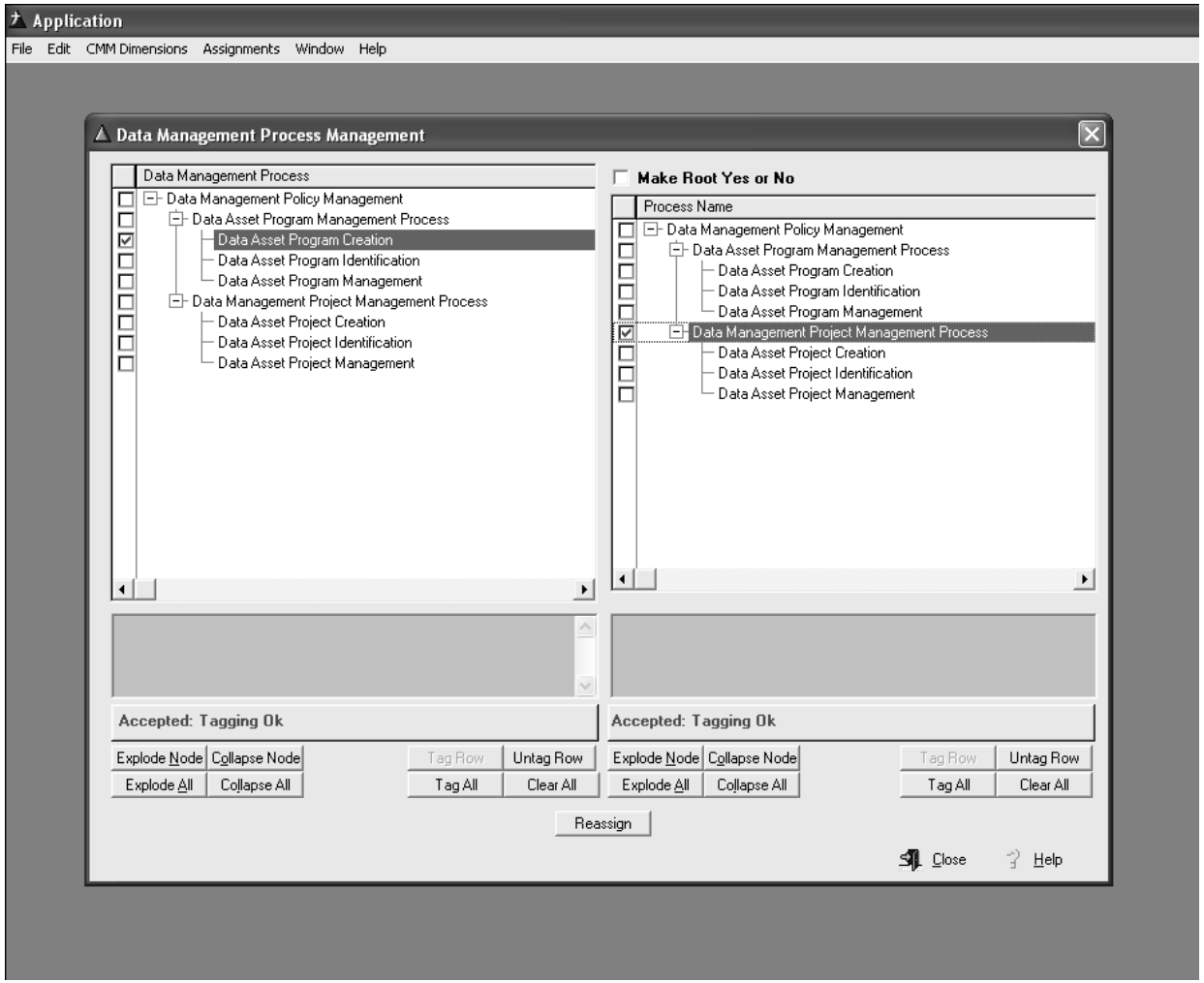
Recursive Dimension and Selected Association Generation



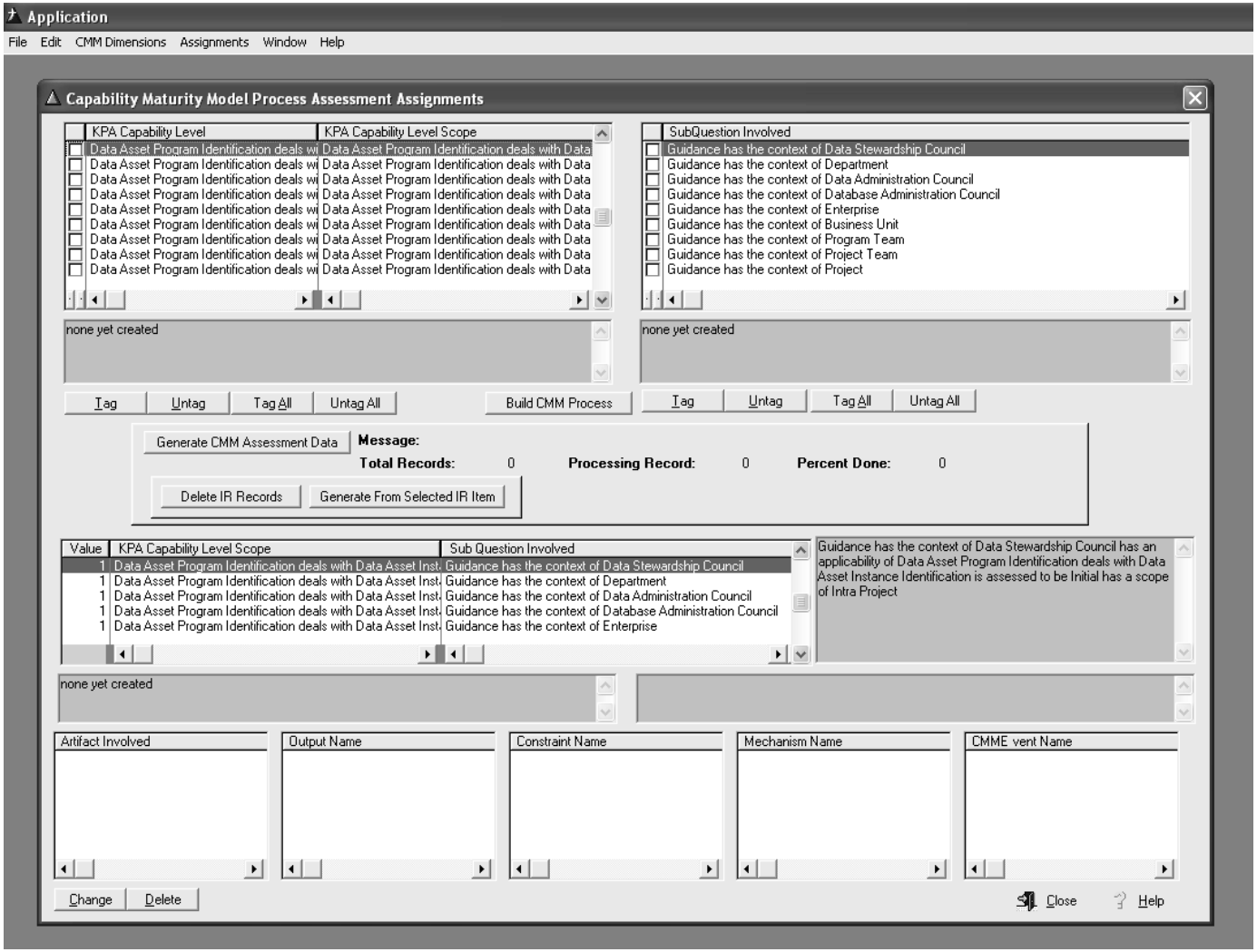
Assignment Processes



ReAssignment Processes



Complete Association Generator Process



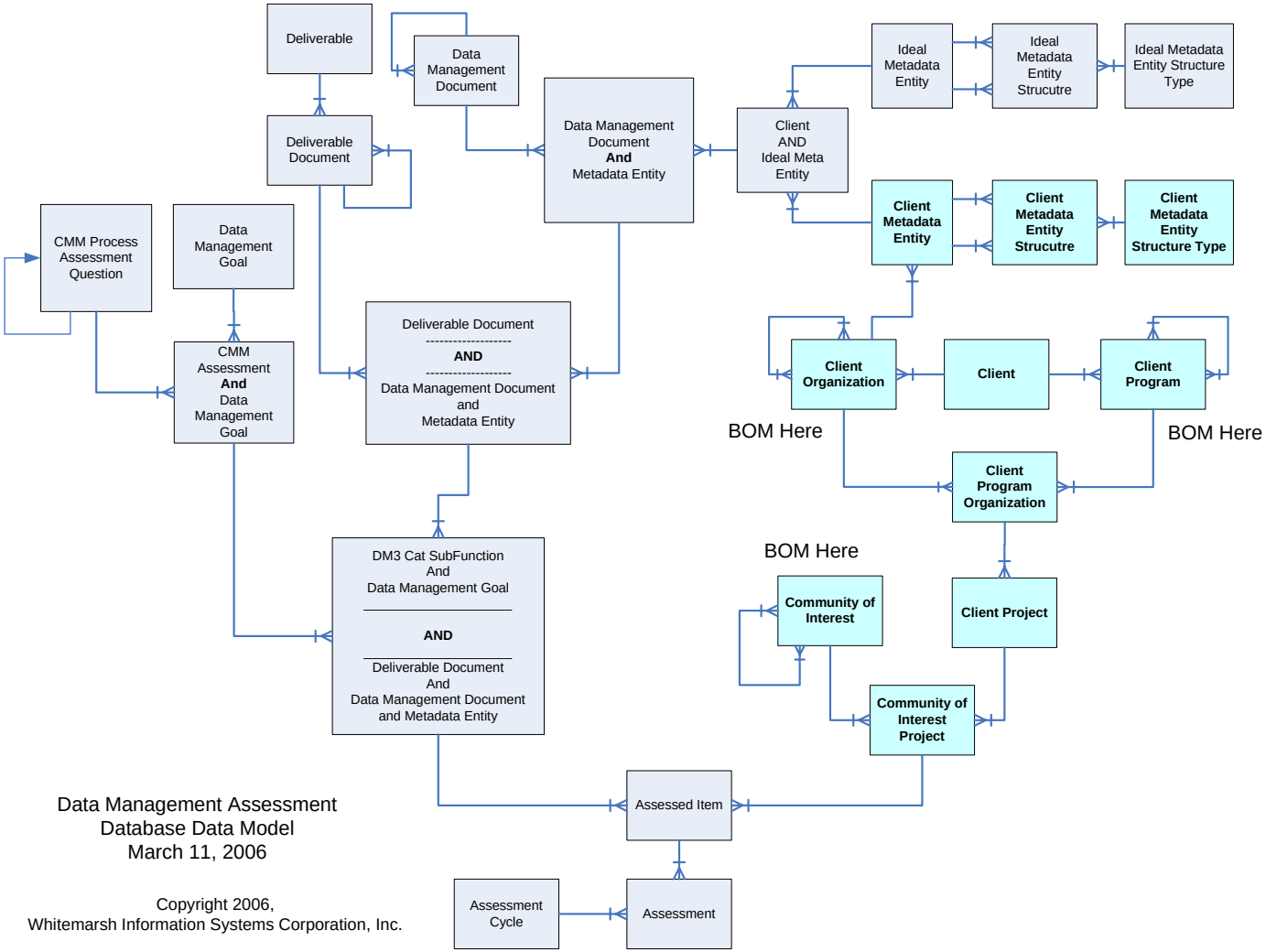
Deletion Processes

- Dimensions
  - ♦ Highlight a dimension row
  - ♦ Press delete
  - ♦ After suitable warnings, CASCADE DELETE over entire association set
- Associations
  - ♦ Highlight an association row
  - ♦ Press delete
  - ♦ After suitable warnings, CASCADE DELETE over entire association set

5.2 Assessor System

- Assessor Data Model Diagram
- Assessor Data Model Tables
- Assessor Process Model

Assessor Data Model Diagram



Assessor Data Model High-Level Tables	
Dimension Table Name	**Description**
Data Management Maturity Model Question	Each row is a question from the DM/CMM Generator System.
Data Management Goals	Each row is a data management goal that is important to the enterprise. E.g., Visibility, Trust, etc.
Data Management Deliverables	Each row represents a data management deliverable that is to be produced during an IT Project
Best Practice Methodology	Each row represents a process in a best-practice methodology that accomplishes either an entire IT project or the data management portion of an IT project. Unit effort metrics are included here too.

Assessor Data Model High-Level Tables	
Dimension Table Name	**Description**
Client Current Practice Methodology	Each row represents a process in the client's current-practice methodology that accomplishes either an entire IT project or the data management portion of an IT project. Unit effort metrics are included here too.
Client Organizations and Communities of Interest	Each row represents a client organization (including for example, functional communities of interest) involved in accomplishing an IT project that involves data management.
Client Programs and Projects	Each row represents a program and/or project that the client is accomplishing that is either an IT project or one that involves data management.

Assessor Data Model High-Level Tables	
Association Table Name	**Description**
DM-CMM question ***VS*** Data Management Goals	Each row represents the association of some data management goal to one or more questions in the DM-CMM questionnaire. This gives you an way to know what processes contribute to which DM Goals.
DM/CMM question and DM Goals ***VS*** Data Management Deliverables	Each row represents the association of a data management goal as assessed by one DM/CMM question is represented by one data management deliverable. This gives you the ability to know "where" to look for the proof of an assessment.

Assessor Data Model High-Level Tables	
Association Table Name	**Description**
Best Practice ***VS*** Client Current Practice	Each row represents the association of Best Practice and Client Current Practice. Left and Right outer joins are created so that the result is a Methodology UNION. Equi-joins represent methodology equivalences. This gives you a comprehensive methodology against which to do an assessment. It additionally gives you the ability to create prescriptions.
Client Organizations ***VS*** Programs and Projects	Each row represents the association of some program or project with a client organization. This gives you the knowledge of which projects and programs within which client organizations to examine and assess.

Assessor Data Model High-Level Tables	
Association Table Name	**Description**
Deliverables ***VS*** Practice	Each row represents the association between a methodology process and a data management deliverable component that is being produced. This gives you the ability to know
Assessments	Each row represents the answer to a specific DM/CMM question in the context of a data management goal that is related to a specific data management deliverable that is accomplished by a methodology that is being performed on a specific projects within the scope of a client organization. Once all these rows are created, the rest is ***"just"*** statistical analysis and prescription writing.

Assessor Process Model

Process Step	Description and Duration Estimate
1.0	Initial contact and interview requirements (1 staff day)
1.1	Identify 3 major programs and organizations. Enter into Assessor database
1.2	Identify 3 major projects in each program. Enter into Assessor database
1.3	Identify key individuals in each program and each project (architect, data management administrator, database administrator, project leaders)
1.4	Sent package of interview product requirements

Process Step	Description and Duration Estimate
2.0	Conduct interviews (3 programs and 9 projects) (12 interviews across 40 staff hours, or five 9-hour days)
2.1	Collect all DM deliverable examples. Enter into Assessor database. Collect the client methodology. Enter into Assessor database. Create association table rows.
2.2	Ask for DM Goals with descriptions. Enter into Assessor database
2.3	Interview clients with respect to all DM/CMM questions

Process Step	Description and Duration Estimate
3.0	Data Analysis and Draft Report Creation (5 staff days)
3.1	Enter all data and generate levels and various materials for the report.
4.0	Visit Client and Present Findings (2 staff days)
4.1	Present findings
4.2	Receive corrections
4.3	Prepare final report and deliver to client
5.0	Quarterly Review for one year (2 staff days)
6.0	Annual (telephone) reassessment for two years (2 staff days)

6.0 Status of the Approach

- DM/CMM Generator System
  - ♦ Data Model Designed and Built into Database Schema
  - ♦ Process Model Designed and built into Working Generator System

- DM/CMM Assessor System
  - ♦ Data Model Designed and Built into Database Schema
  - ♦ Process Model Designed and built into 1st Version Generator System

7.0 Practical Benefits of a DM/CMM

Data Driven vs Process Driven

Quantity of Tables From estimate of a prototypical database	Average Columns per Table	Total Columns	Process Driven Approach (2 hours per column)	Data Driven Approach (2 hours per table, plus 1/30th 11179 Data Elements	
400	15	6,000	5.75 staff years	0.7 Staff years	**Cost Difference at $100 per hour**
		Cost	$1,200,000	$120,000	$1,180,000 in favor of data driven.

Re-Use Based Data Management

Activity	Quantity	Cost via technique employed for definition
Starting quantity of columns/fields	19,000	$6.75 million
Elimination of closely named columns and fields reduced the quantity to	3,000	$1.06 million
Elimination of same concept but very differently named columns and fields reduced the quantity to	560	$200,000