At the Intersection of Data Governance and Data Quality

Metadata Intelligence

Don Soulsby
Sandhill Consultants
Sandhill Consultants
Strengths are in our technology understanding of foundational solutions such as Modeling, Metadata, Data Quality and Data Governance as well as data operations the their supported platforms. We offer sales, services, education and support, as well as standards and best practices should clients need these services.

Don Soulsby
• Vice President Architecture Strategies, Sandhill Consultants
• Enterprise Architecture, Business Intelligence, Metadata Consultant
• Data/Process modeller
• Educator
• Industry Speaker

metaWright
The New Cx Executives

Ceo – Chief eCommerce

CIO – Chief Internet Officer

CDO – Chief Digital Officer

CDO – Chief Data Officer

Organizations, Processes Come and Go, Data Persists
The Corporate Brain Challenge

SOFTWARE
Programming Language

DATABASE

WETWARE
Human Language

TAXONOMY
Manage Dis!
You don’t have a **filing** problem,

You have a **finding** problem”
Ownership

This is a story about four People named EVERYBODY, SOMEBODY, ANYBODY and NOBODY. There was an important job to be done and EVERYBODY was sure that SOMEBODY would do it. ANYBODY could have done it, but NOBODY did. SOMEBODY got angry about that, because it was EVERYBODY’s job. EVERYBODY thought ANYBODY Could do it, but NOBODY realized that EVERYBODY wouldn’t do it. It ended up that EVERYBODY blamed SOMEBODY when NOBODY did what ANYBODY could have.
Data Ownership

That which is Everybody’s business is Nobody’s business
Izaak Walton

Everybody’s business is Nobody’s business, and Nobody’d business is My business
Clara Barton
DAMA – Body of Knowledge

- Data Architecture Management
- Data Development
- Database Operations Management
- Data Security Management
- Reference & Master Data Management
- Data Warehousing & Business Intelligence Management
- Document & Content Management
- Meta Data Management
- Data Quality Management
- Data Governance
The Capability Maturity Model

Initial
(1)  
ad hoc Process

Repeatable
(2)  
Stable Process

Defined
(3)  
Standard Process

Managed
(4)  
Measured Process

Optimizing
(5)  
Improving Process

Chaotic

Controlled Environment

Consistent Execution

Quality and Productivity Improvement

Continuing Improvement

Basic Management Control

Process Definition

Process Measurement

Process Control
CMM History

• The Software Engineering Institute established 1984 at Carnegie Mellon University, Pittsburgh
• Principle customer is the Department of Defense
• U.S. Air Force Project 1987
  o Method to select software contractors
• Over 5,000 Assessments performed since 1987
Process Improvement Models

• 120+ and counting
• Data MM
• Training MM
• Release MM
• Configuration MM
• Documentation MM
• Business Rules MM
• Etc...

Beyond The CMM:
Brett Champlin, 2002
Data Related Maturity Models

- Data Categories MM (BC Ministry of Forests)
- Data MMM (MITRE)
- Data MMM (Giga)
- Data Resource Management MM (Champlin)
- Data Warehouse Information Management MM (Ladley)
- Data Warehousing MM (Marco)
- Enterprise-wide Data MMM (Parker)
- Information Delivery MM (Computer Associates)
- Information Evolution Model (SAS)
- Information Maturity Framework (BC Ministry of Transp)
- Information Quality MM (English)
- Total data Quality Management MM (MIT)
- Stages of Growth (Nolan)

Beyond The CMM:
Brett Champlin, 2004
• About 1995, a team led by Burt Parker
• Basically changed the word “software” to “data” for the KPAs
• Didn’t define Goals and Activities
• But, they looked at the larger context
  – Information Management
  – DM Purpose Model
  – DM Normative Model
  – DM Utility Model
  – DM Roles

Beyond The CMM:
Brett Champlin, 2002
Measuring Data Management Practice Maturity: A Community’s Self-Assessment

Peter Aiken, Virginia Commonwealth University/Institute for Data Research
M. David Allen, Data Blueprint
Burt Parker, Independent consultant
Angela Mattia, J. Sergeant Reynolds Community College
Maturity Model Adoption Curve – Six Stages

1. Prenatal
2. Infant
3. Child
4. Teenager
5. Adult

Business Value
Data Consolidation
Semantic Integration

Spreadsmarts
"Production Reporting"
"Data Marts"
"Data Warehouses"
"Enterprise DW"

The Data Warehouse Institute

1998

CROSSING THE CHASM
A BUSINESSWEEK BESTSELLER
MARKETING AND SELLING DISRUPTIVE PRODUCTS TO MAINSTREAM CUSTOMERS

GEOFFREY A. MOORE

SANDHILL
Der Reifegrad des Datenmanagements

Entwicklungsstufen

Ziele
- Information als Unternehmensresource
- Eigene Zuständigkeit für Daten
- Datenstandardisierung
- Einsatz von DBMS
- Isolierte Anwendungen
- Funktionsorientierung

Treiber
- Information als betriebliche Ressource
- Prozessorientierung
- Integrationsanforderungen
- Technologie

Jahr
- 1960
- 1970
- 1980
- 1990
- 2000

In Anlehnung an E. Ortner 1991

Beyond The CMM:, Brett Champlin, 2004
Yes, even your humble presenter

DATA ANALYTICS MM

Level 1 - Initial
- Programmers
- Extract files, no metadata
- Data access tools
- Data movement tools
- Personal Data

Level 2 - Repeatable
- Project Management, Standards
- DBA’s, Modelers, Specialists
- Reporting Tools
- Data quality
- Team Data

Level 3 - Defined
- Metadata Management
- Architects
- Content Management
- Extended Team Data

Level 4 - Managed
- Measurement
- Integrated Content Management
- Corporate Data

Level 5 - Optimizing
- Feedback System
- Knowledge Management
- Extended Enterprise Data
Management’s Role

Executive Maturity Must Precede Process Maturity

Continuously Improving Process

Optimizing (5)

Managed (4)

Defined (3)

Repeatable (2)

Initial (1)

Recognized (2)

Supported (3)

Participative (4)

Managing (5)

Need Executive Oversight

Need Management Program

Need Consistency

Continuously Improving Process

Continuously Improving Process

Predictable Process

Standard, Consistent Process

Disciplined Procedures

CMMI Institute, powered by Carnegie Mellon
The CMMI Institute is a Carnegie Mellon corporation, the new home of the Capability Maturity Model Integration (CMMI)
Formerly managed by the Software Engineering Institute (SEI)
CMMI for Development, CMMI for Services
CMMI Assessment

3.1415
The DMM model, first announced in May 2014, was developed using the principles and structure of CMMI Institute’s Capability Maturity Model Integration (CMMI®)

- Data Management Strategy
- Data Governance
- Data Quality
- Data Operations
- Platform & Architecture
- Supporting Processes
DMM Assessment Process

- Prioritize Practice Areas
- Communicate Goals
- Respond to Core Questions
- Establish Practice Area maturity levels
  - Not started, Partially complete, Complete
- Develop People, Process & Product recommendations to advance to next level
DMM – Practice Areas

Data Management Strategy
- Data Management Strategy
- Communications
- Data Management Function
  - Business Case
  - Funding

Data Governance
- Governance Management
- Business Glossary
- Metadata Management

Data Quality
- DQ Strategy
- Data Profiling
- DQ Assessment
- Data Cleansing

Data Operations
- Data Requirements Definition
- Data Lifecycle Management
- Provider Management

Platform & Architecture
- Architectural Approach
- Architectural Standards
- Data Management Platform
- Data Integration
- Historical Data, Archiving
Data Management Maturity Assessment

POWER
OF
3

SANDHILL
EDME
EDME Certification:
• Fully qualified to lead a DMM Assessment and provide high value to any organization
• Fully qualified to lead data management improvement programs
Anonymous Benchmark
Data Governance Capabilities

DATA GOVERNANCE

Governance Management
Business Glossary
Metadata Management

CMMI Institute

SANDHILL
Data Governance

• **Governance Management**
  Develop the ownership, stewardship, and operational structure needed to ensure that corporate data is managed as a critical asset and implemented in an effective and sustainable manner.

• **Business Glossary**
  Supports a common understanding of terms and definitions about structured and unstructured data supporting business processes for all stakeholders.

• **Metadata Management**
  Establish the processes and infrastructure for specifying and extending clear and organized information about the *data assets under management.*
Transformation NOT Translation

Unambiguous Business Terminology

Business Glossary

Conceptual Data Model

Logical Data Model

Data Delivery Model
“Content is King,
Context is the Crown”

Eric Schmidt, former CEO at Google

Not Data about Data,
Data in Context
The Burnt Stick in Context

- Fish Bones = FUEL
- Cave Drawing = PENCIL
The Grid

- Important to understand the relationship of artifacts to one another by proximity.
- Important to understand the precedence of artifacts by depth.

...Carnegie Museum of Natural History in Pittsburgh that researchers said was "mixed up" with another dinosaur skull shortly after they were discovered in Utah in 1909.

YALE BRONTOSAURUS GETS HEAD ON RIGHT AT LAST
Special to the New York Times - October 26, 1981
From the Latin "meta", which means heap, pile... assuming the meaning of high ground.
Data Quality Capabilities

DATA QUALITY

- Data Quality Strategy
- Data Profiling
- Data Quality Assessment
- Data Cleansing

SANDHILL

CMMI Institute
Data Quality

- **Data Quality Strategy**
  Defines an integrated, organization-wide strategy to achieve and maintain the level of data quality required to support the business goals and objectives.

- **Data Profiling**
  Develops an understanding of the content, quality, and rules of a specified set of data under

- **Data Quality Assessment**
  Provides a systematic approach to measure and evaluate data quality according to processes, techniques, and against data quality rules.

- **Data Cleansing**
  Defines the mechanisms, rules, processes, and methods used to validate and correct data according to predefined business rule
Data Quality Issues

SYS A  SYS B  SYS C

BI SYSTEM
Some More Cases to Consider

• 75% of companies report significant problems due to defective data
• 92% of claims Medicare paid to community health centers over one year’s time were “improper or highly questionable”
• Wrong price data in retail databases costs consumers as much as $2.5 billion in overcharges annually
• 96,000 IRS tax refund checks were returned as undeliverable one year
Data Profiling

- **Accuracy** – criteria related to affinity with original intent, veracity as compared to an authoritative source, and measurement precision
- **Completeness** – criteria related to the availability of required data attributes
- **Coverage** – criteria related to the availability of required data records
- **Conformity** – criteria related to alignment of content with required standards
- **Consistency** – criteria related to compliance with required patterns and uniformity rules
- **Duplication** – criteria related to redundancy of records or attributes
- **Integrity** – criteria related to the accuracy of data relationships (parent and child linkage)
- **Timeliness** – criteria related to the currency of content and availability to be used when needed
Coverage & Integrity
Conformity & Duplication
Timely

Data Quality Scorecard

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>A</th>
<th>T</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>CRM</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>
Data Quality Assessment

"You can't manage, what you can't measure"
"You can't measure, what you can't find!"

Measure = Observation

Sandhill Enterprise Modelling Set of Standards
Data Management

DATA QUALITY

DATA GOVERNANCE

METADATA INTELLIGENCE
Many to Many Problem

Many Sources

Many Uses
Many to Many Solution

Many Sources

Enterprise Data Warehouse

Many Uses
Business Intelligence

Big Data

Data Bases

ERP

ETL

Data Warehouse

SANDHILL
Data Asset Inventory

Self Service BI
Value of Metadata

What does it mean?

UNDERSTANDIBILITY
Synonyms, Homonyms, Anti-nyms

BUSINESS GLOSSARY

Where does it come from?

TRACEABILITY

METADATA REPOSITORY
Business Concept to Database
WHY?
The Issue

Do Executives Trust Data or Intuition in Decision-Making?

By Bob Violino
JUN 9, 2014 3:44pm ET

“Trust yes, but verify.”

Compliance
ALL DATA, Great and Small

SANDHILL
TRUST

Firm reliance on the integrity, ability, or character of a person or thing

Custody; Stewardship

Something committed into the care of another; charge.
Meatadata Traceability

We can trace it, so you can trust it.

All our fresh beef is only ever sourced from the UK, from specific breeds and carefully selected farms. Which is why you can be sure that every cut of Marks & Spencer beef will be a choice one.

MARKS & SPENCER

EXCLUSIVELY FOR EVERYONE

www.marksandspencer.com
FROM FIN TO FORK

1. Code the Catch
   Fish harvesters identify their catch with a unique code using a tag or label.

2. Upload Info
   Info about when, where and how your seafood was caught is uploaded to ThisFish.info.

3. Handle + Ship
   Other folks in the supply chain can also upload info about its handling and shipping.

4. Trace to Discover
   Trace your seafood’s code at ThisFish.info to discover its story and connect to your harvester.

SANDHILL
Data Quality going Forward

OBSERVATIONAL Analytics

ALGORYTMIC Analytics

SANDHILL
The Only Thing That Is Constant Is Change*

* Heraclitus – 500 BCE
Model Driven Change Management

- Business Domain
- Technology Domain
- Logical Data Model
- Conceptual Data Model
- Forward Engineering
- Reverse Engineering
- Middle Out
- Business Glossary
- Human Oriented
- Computer Oriented
- PROACTIVE
- REACTIVE
- Metadata Repository
- DATABASE
- SANDHILL
Not Just DATA

DATA

- Conceptual Data Model
- Logical Data Model
- Data Delivery Model

PROCESS

- Business Process Model
- System Design Model
- Application Model

Business Change

Application Database Change

SANDHILL
More than Data Metadata

Scanner

Big Data
Erwin Models
Data Bases
Applications
ERP

Metadata Repository

SANDHILL
Karl Ludwig von Bertalanffy was an Austrian-born biologist known as one of the founders of general systems theory (GST).

IDEF-0*

inputs → activity → outputs
controls
mechanisms

*ICAM Definition, Integration DEFinition
Integrated Computer-Aided Manufacturing
1. What is was When it Was
2. What it Was as if it Was Now
"... please which way I ought to go from here?"
said Alice.
"That depends a good deal on where you want to get to," said the Cat.
"I don't much care where----" said Alice.
"Then it doesn't matter which way you go," said the Cat.
Data Governance Mall

Mall Taxonomy

Where am I?

“WayFinding”

You are here

Nice to know how to get there from here!
Fit For Use – by the Consumer


SANDHILL
Sandhill Consultants:

Sandhill Consultants
Is a group of people, products and processes that help our clients:

• Build comprehensive **Data Architectures**
• Resulting from a persistent **Data Management Process**
• Founded on a robust **Data Governance Practice**
• With Da’tum Ver’it’as – **Trusted, Verifiable Data**
Tibetan Proverb

“If upstream is dirty, downstream will be muddy”

Thank You

Don.Soulsby@SandhillConsultants.com

[shuǐtóu] [bù] [qīngshuǐ] [wěi] [hún]

SANDHILL