Successful Metadata Management Strategy and Implementation for State Government

By David Marco
President
EWSolutions

November 30, 2017

Sponsored by the NASCIO Data Management Working Group

EWSolutions' Background

EWSolutions is a Chicago-headquartered strategic partner and full life-cycle systems integrator providing both **award winning** strategic consulting and **full-service implementation services**. This combination affords our client partners a full range of services fo any size enterprise information management, metadata management, data governance and data warehouse/business intelligence initiative. Our notable client partner projects have been featured in the Chicago Tribune, Federal Computer Weekly, Journal of the American Medical Informatics Association (JAMIA), Crain's Chicago Business, The Doingsand won the 2004 Intelligent Enterprise's RealWare award, 2007 Excellence in Information Integrity Award nomination, DM Review's 2005 World Class Solutions award and 2016 CIO Review 20 Most Promising Enterprise Architecture providers.

CorporateLiveWire INNOVATION & EXCELLENCE AWARDS 2018 — WINNER —

Most Innovative in Data Management Solutions

Over 100 Client Partners

100% Data Management Success since 1997

Chicago Tribune

Information Integrity Coalition

2007 Excellence in Information Integrity Award Nomination

JAMIA

20 MOST PROMISING ENTERPRISE ARCHITECTURE TECHNOLOGY PROVIDERS - 2016

FCW THE BUSINESS OF FEDERAL TECHNOLOGY

DMReview 2005

World Class Solutions Award Data Management

intelligent enterprise REAL Awards TRANSFORM 2004 WINNER

Best Business Intelligence Application Information Integration Client: Department of Defense

DMF DATA MANAGEMENT FORUM

Excellence and Thought Leadership

CRAIN'S CHICAGO BUSINESS

Federal Bureau of Investigation
Meritorious Service Award

DePaul University

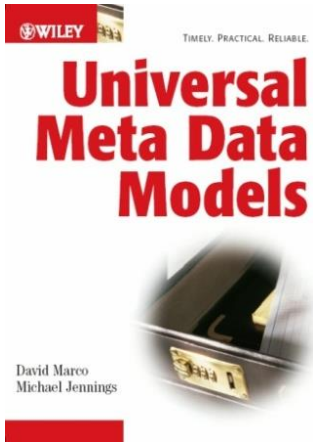
NASCIO

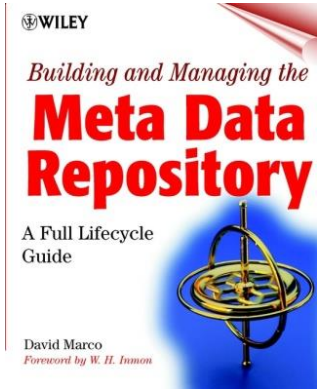
For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, email us at info@EWSolutions.com

www.EWSolutions.com

David Marco – Professional Profile

Best known as the world's foremost authority on metadata management and the father of the Managed Metadata Environment, he is an internationally recognized expert in the fields of data governance, big data, data warehousing, master data management and data management. In 2004 David Marco was named the **"Melvil Dewey of Metadata"** by **Crain's Chicago Business** as he was selected to their very prestigious **"Top 40 Under 40"** list. David Marco has authored several books including the widely acclaimed **"Universal Metadata Models"** (Wiley, 2004) and the classic **"Building and Managing the Metadata Repository: A Full Life-Cycle Guide"** (Wiley, 2000).



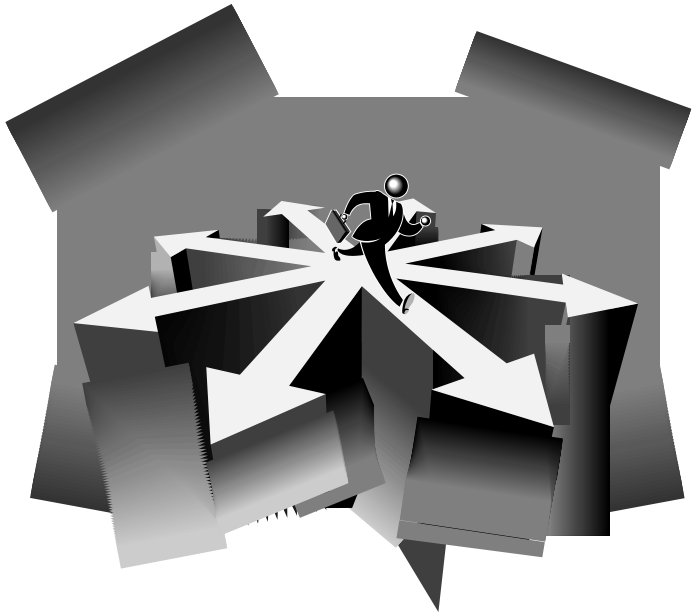


- President of Data Management University (**DataManagementU.com**)
- Author of several best selling information technology books, including the top 2 sellers in metadata management
- **2017** Sabi University, Academic Board member
- **2016** Channel Expert for Business Analytics Collaborative
- **2008 DAMA Data Management Hall of Fame (**Professional Achievement Award**)**
- **2007 DePaul University named** him one of their **"Top 14 Alumni Under 40"**
- Selected to the prestigious **2004 Crain's Chicago Business "Top 40 Under 40"**
- Presented hundreds of keynotes/seminars across four continents
- Published hundreds of IT articles some of which were translated into Mandarin, Russian, Italian, Portuguese and others
- Taught at the **University of Chicago** and **DePaul University**
- Holds CDMP, CDP, CCP and CBIP certifications

Email: DMarco@EWSolutions.com

Agenda

- Metadata Management Fundamentals
- Business vs. Technical Metadata
- Metadata Management Use Cases
- Metadata ROI
- Metadata Management and Data Management



Metadata Management Fundamentals

What is Metadata?

Metadata

By definition metadata is

1. “Data about data”.
2. “Everything that data is not”.

What is Metadata?

Metadata Definition

Metadata is a type of data that digitally describes the who, what, when, where, why, and how of an organization's data, processes, applications, assets, business concepts, and/or other things of interest.

Metadata Is Knowledge

Metadata vs. Data

- ***Metadata:*** Metadata contains the knowledge that a **1)** field is called "Customer_Name", is 40 characters in length, and exists in systems A, B, and C; **2)** that our company has 3 systems which contain customer master data. These systems are…
- ***Data:*** Data would be a specific instance of "Customer_Name" equaling "John Doe"
- ***Information:*** Data that is meaningful to a business user. They understand it and they know what to do with it

Metadata Management Fundamentals

Information = Data + Metadata

(content) *(context)*

Managed Metadata Environment Return on Investment (ROI)

Managed Metadata Environment ROI

“The key to your company’s prosperity is how well you gather, retain and disseminate knowledge”

“Managed metadata environments are the key to gathering, retaining and disseminating knowledge”

“How can our enterprise be agile when we don’t even know the who, what, when, where, how and why of our data?”

Managed Metadata Environment ROI

- Metadata for the Business (business metadata)

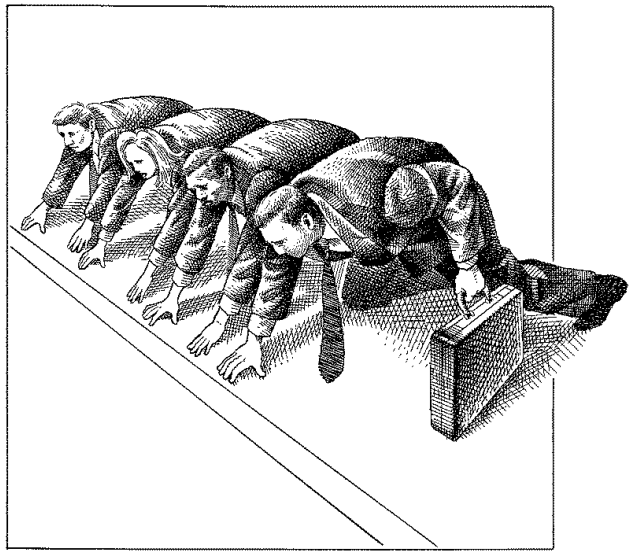
- Metadata for the IT Department (technical metadata)

Business Metadata

Managed Metadata Environment ROI

Metadata for the Business (business metadata)

- Provides the semantic layer between a company's systems (operational and business intelligence) and their business users



Metadata for the Business

- Reduces training costs
- Makes strategic information (e.g. data warehousing, customer relationship management (CRM), supply chain management (SCM), enterprise resource planning (ERP), etc.) much more valuable as it aids analysts in making more profitable decisions
- Create **actionable** information
- Limits incorrect decisions
- Assists business analysts in finding the information they need, in a timely manner
- Bridges the gap between business users and IT professionals
- Increases confidence in the IT system data
- A **MUST** for Data Governance

Business Metadata In Action

Metadata for the Business

2017 Monthly Global Sales Report				February 7, 2017
Month	**Product Category**	**Sales $ (in thousands) U.S**	**Sales $ (in thousands International**	**Sales $ (in thousands Total**
December	TV	22,101	10,200	32,301
	DVD	11,190	4,300	15,490
	Cellular Phone	12,190	7,193	19,383
	Digital	4,002	1,301	5,303
	Miscellaneous	1,209	870	2,079
November	TV	42,000	22,200	64,200
	DVD	21,190	9,878	31,068
	Cellular Phone	28,193	12,193	40,386
	Digital	8,901	901	11,802
	Miscellaneous	2,730		4,260
October	TV	70,100	32,	103,050
	DVD	31,900	14,87	46,778
	Cellular Phone	41,700	17,550	59,250
	Digital	20,000	4,100	24,100
	Miscellaneous	4,850	2,850	7,700

"Sales $ U.S." is comprised of aggregated sales revenues from the United States, Canada, and Mexico, but does not subtract sales dollars from returned orders

Metadata Makes for Better Decisions

Metadata for the Business

2017 Monthly Global Sales Report				February 7, 2017
Month	**Product Category**	**Sales $ (in thousands) U.S**	**Sales $ (in thousands International**	**Sales $ (in thousands Total**
December	TV	22,101	10,200	32,301
	DVD	11,190	4,300	15,490
	Cellular Phone	12,190	7,193	19,383
	Digital	4,002	1,301	5,303
	Miscellaneous	1,209	870	2,079
November	TV	42,000	22,200	64,200
	DVD	21,190	9,878	31,068
	Cellular Phone	28,193	12,193	40,386
	Digital	8,901	2,901	11,802
	Miscellaneous	2,730	1,530	4,260
October	TV	70,100	2,950	103,050
	DVD	31,900	78	46,778
	Cellular Phone	41,700		59,250
	Digital	20,000	4	24,100
	Miscellaneous	4,850	2,85	7,700

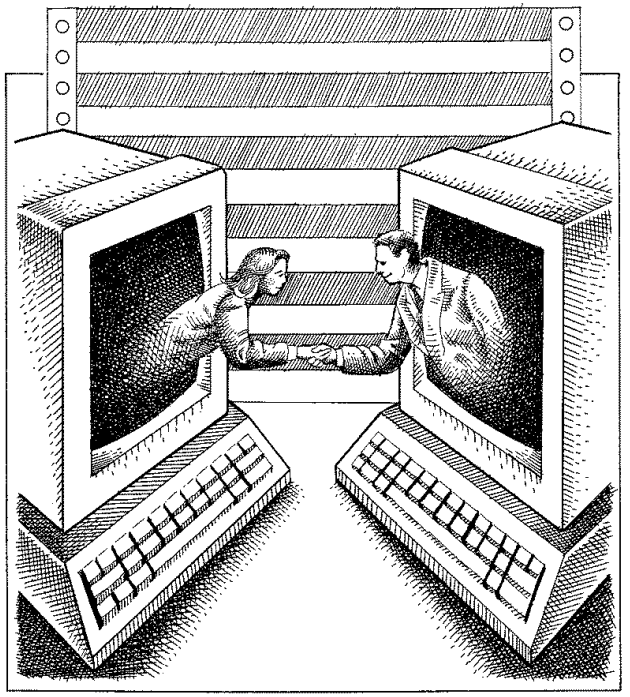
Information Quality Tracking Statistics
8.4% of the dollar values were not loaded
1.7% of the records were not loaded

Technical Metadata

Managed Metadata Environment ROI

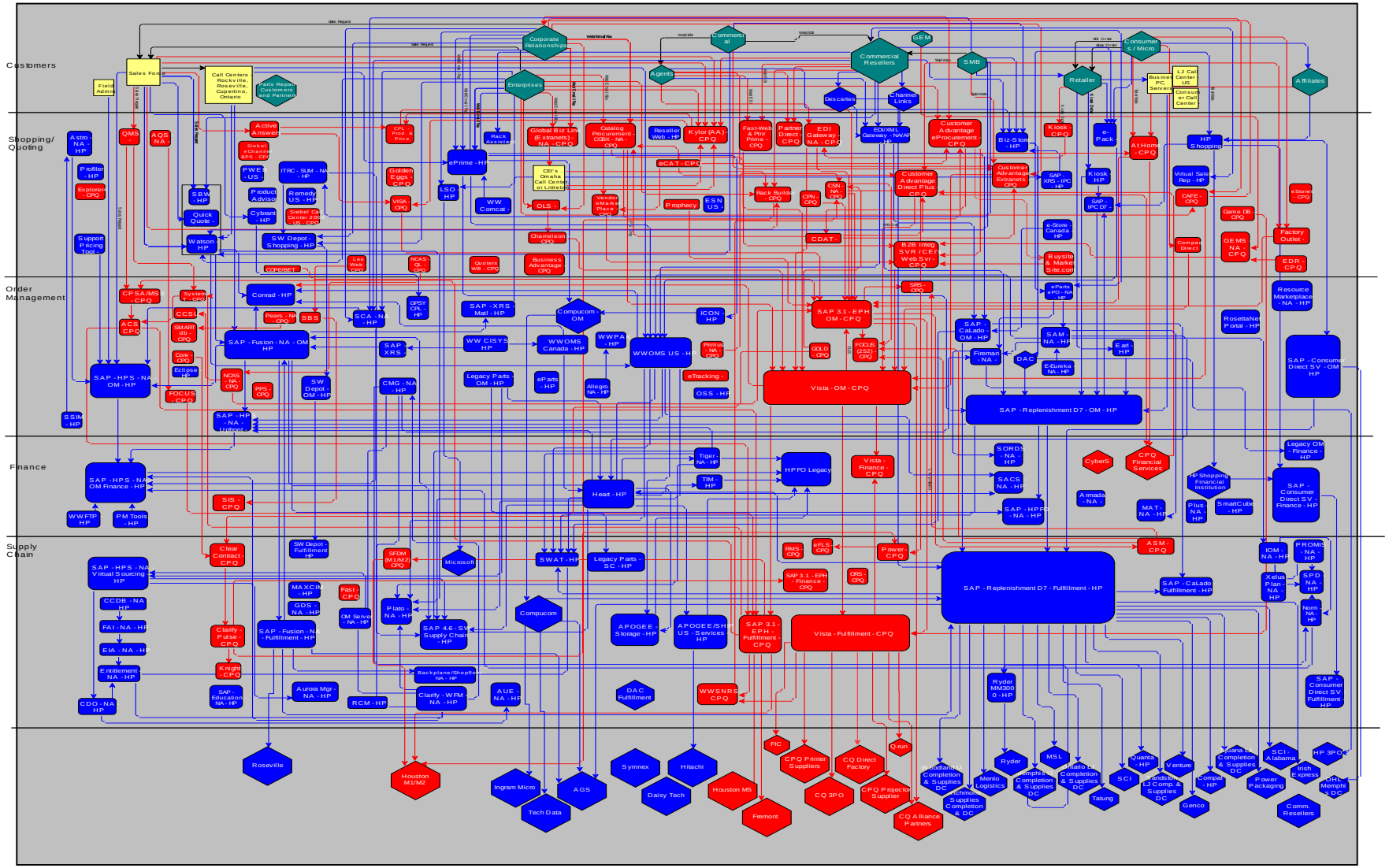
Metadata for the IT (Information Technology) Department is Technical Metadata

- Help IT departments better manage, maintain and grow their IT systems and assets



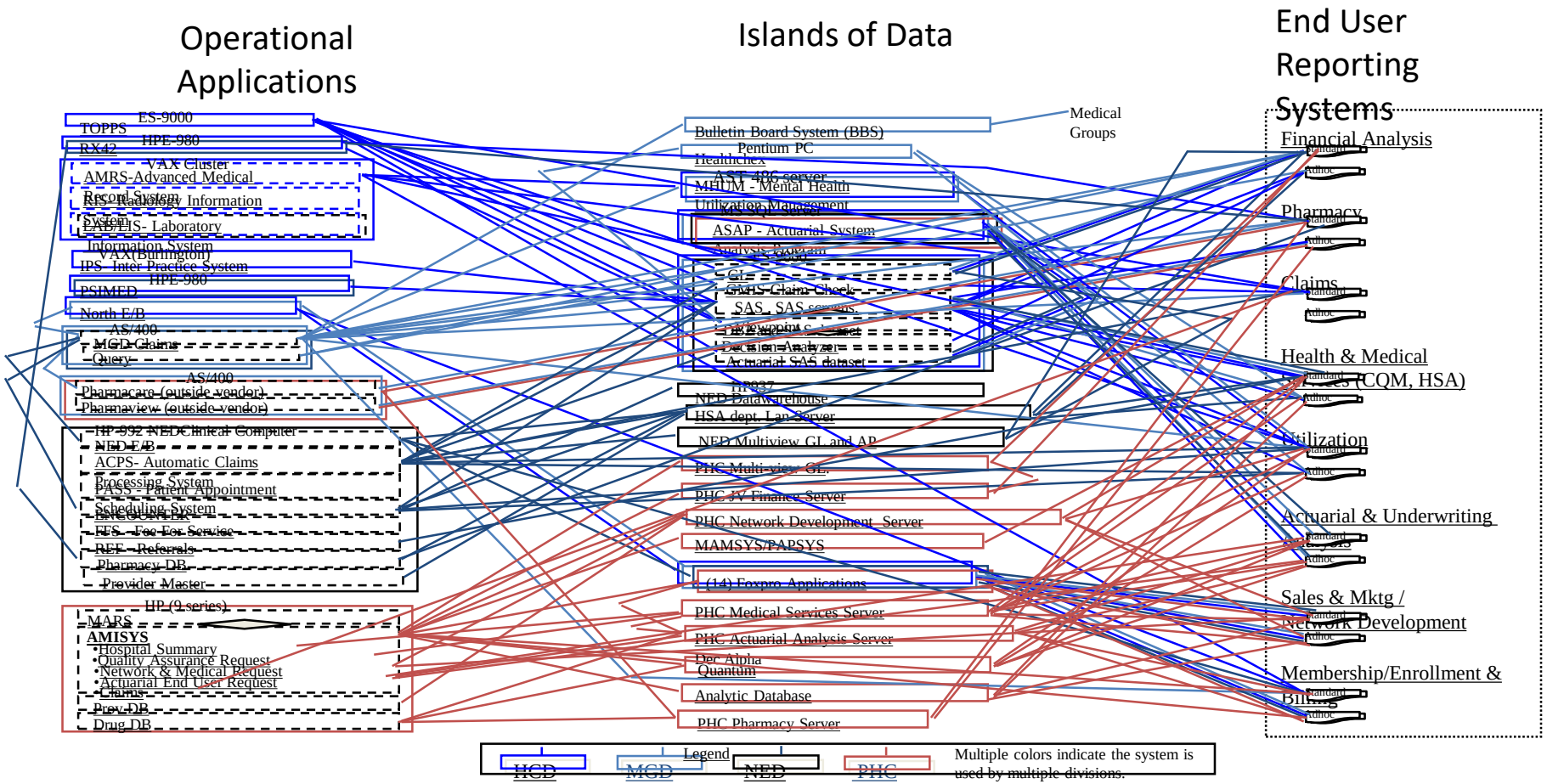
What Does One Process Look Like for a Large Company?

Fortune 100 - One Process



What Do Our Current IT Systems Look Like?

Islands of Data



Metadata for the IT Department

- **Dramatically reduces the probability of project failure**
- Speeds system's time-to-market
- Reduce system development life-cycle time
- Limit redundant data
- Limit redundant processes
- Managing IT portfolios
- Leverage work done by other teams
- Reduced rework
- Reduce research time
- Reduce unproductive work
- Lowers the impact of staff turnover

Technical Impact Analysis

Metadata for the IT Department

Question: Show all decision support tables/files, programs, and fields impacted by a change to the "CUST" table in the "Order Entry" system

Impact Analysis Report						January 7, 2016
Source System	**Source Table**	**Impact Field**	**Program Impacted**	**Tables/Files Impacted**	**Table Type**	**Fields Impacted**
Order Entry	CUST	Customer_Name	CUSTOMER_PR02	DW_CUSTOMER	T	Cust_Name_First
						Cust_Name_Middle
						Cust_Name_Last
			CUSTOMER_PR01	I02_CUSTOMER	I	Cust_Name_First
						Cust_Name_Middle
						Cust_Name_Last
		Customer_Addr	CUSTOMER_PR02	DW_CUSTOMER	T	Cust_Name_Address
						Cust_Name_City
						Cust_Name_State
						Cust_Name_Zip
			CUSTOMER_PR01	I02_CUSTOMER	I	Cust_Name_Address
						Cust_Name_City
						Cust_Name_State
						Cust_Name_Zip

***Legend**

"T" = Target

"I" = Intermediate

"S" = Source

Metadata for the IT Department

Question: Show all systems, tables/files, fields, and their domains impacted by a change to the length of all occurrences of the Customer_Name field

Impact Analysis Report				January 7, 2016
Field	**System**	**Tables/Files**	**Fields**	**Domain**
Customer Name	Order Entry	CUSTOMER_BILL_TO	CUST_NAME	Alphanumeric 20
		CUSTOMER_SELL_TO	CUST_NAME	Alphanumeric 20
		CUSTOMER_SHIP_TO	CUST_NAME	Alphanumeric 20
		ORDER_HEADER	CUST_NAME	Alphanumeric 20
		ORDER_DETAIL	CUST_NAME	Alphanumeric 20
	General Ledger	CUSTOMER	Cust_Name	Alphanumeric 35
		EXPENSES	Cust_Name	Alphanumeric 35
		CUST_ACCOUNTS	Cust_Name	Alphanumeric 35
	Data Warehouse	DW_CUSTOMER	Cust_Name	Alphanumeric 20
		I01_CUSTOMER	Cust_Name	Alphanumeric 20
		I02_CUSTOMER	Cust_Name	Alphanumeric 20
		I03_CUSTOMER	Cust_Name	Alphanumeric 20
	Data Mart - Marketing	DM_CUSTOMER	DM_Cust_Name	Alphanumeric 20
		I01_DM_CUSTOMER	DM_Cust_Name	Alphanumeric 20
		I02_DM_CUSTOMER	DM_Cust_Name	Alphanumeric 20

MME For Systems Consolidation

Metadata for the IT Department

Systems Consolidation Report		BigCity Bank		August 15, 2017			
BigCity Bank				Small Town Bank			
Attribute Name	Attribute Definition	Entity Name	System Name	Attribute Name	Attribute Definition	Entity Name	System Name
Cust_Nbr	Cust_Nbr is the attribute of record for BigCity Bank customer numbers	Cust_Tbl	Central Customer System	CUSTNUM	Customer numbers from the deposit system.	CUSTTABLE	CUSTAPPL
				Purchase_No	Customer numbers from the purchase in the legacy deposit system	Purch_Tbl	CUSTSYS
				Borwr_No	Customer numbers from the loan system.	Borrower_File	LoanSys
Cust_Type	Cust_Type is the attribute of record for BigCity Bank customer types (affluent, upward, standard, high risk).	Cust_Tbl	Central Customer System	CUSTCDE	Customer types from the general ledger system.	GL_CUST	GLAPPL
Cust_Card_Ind	Cust_Card_Ind is the attribute of record for BigCity Bank customer 's that have a BCB credit card.	Cust_Tbl	Central Customer System		None applicable		
Cust_Crdt_Ratg	Cust_Crdt_Ratg is the attribute of record for BigCity Bank customer credit ratings (Superior Risk, Low Risk, Standard Risk, High Risk, Extreme Risk).	Cust_Tbl	Central Customer System	Credit_Rate	Customer rate is from the general ledger system and refers to the credit rating/worthiness of a customer.	GL_CUST	GLAPPL

Metadata for the IT Department

Systems Consolidation Report				BigCity Bank				August 15, 2017
BigCity Bank					Small Town Bank			
Entity Name	Attribute Definition	Attribute Name	Domain Value	Transformation Rules	Attribute Name	Domain Value	Entity Name	
Cust_Tbl	Cust_Type is the attribute of record for BigCity Bank customer types: 1 = affluent 2 = upward 3 = standard 4 = high risk	Cust_Type						
			1	Cust_Type = 1 WHEN CUSTCDE = 3 AND CUSTBAL > 500,000	CUSTCDE	3	GL_CUST	
					CUSTBAL	High cardinality field	GL_CUST	
			2	Cust_Type = 2 WHEN CUSTCDE = 4 AND CUSTBAL <= 500,000 AND CUSTBAL > 200,000	CUSTCDE	3	GL_CUST	
					CUSTBAL	High cardinality field	GL_CUST	
			3	Cust_Type = 3 WHEN CUSTCDE = 1 or 2 AND CUSTBAL <= 200,000 AND CUSTBAL > 75,000	CUSTCDE	3	GL_CUST	
					CUSTBAL	High cardinality field	GL_CUST	
			4	Cust_Type = 4 WHEN CUSTCDE = 0 AND CUSTBAL < 75,000 AND Credit_Rate < 22	CUSTCDE	3	GL_CUST	
					CUSTBAL	High cardinality field	GL_CUST	
					Credit_Rate	High cardinality field	GL_CUST	
Cust_Card_Ind	Cust_Card_Ind is the attribute of record for BigCity Bank customer 's that have a BCB credit card.	Cust_Tbl						

The Cost of Redundancy

- Large healthcare insurance company
- Has a $1.6 billion IT budget
- They estimate it costs them $2 per month to store each gigabyte of data
- $8 per month if you add in services and maintenance
- They estimate that they have 1.6 petabytes of redundant data
- What does this cost them yearly? Simple math
- $8 x 12 months x 1,000,000 (1.6 petabytes) = **$153,600,000**

Managed Metadata Environment ROI

"We Build Systems To Manage Every Aspect Of Our Business, Except One To Manage The Systems Themselves."

"A Managed Metadata Environment Is A System That Manages Our Systems."

Metadata Management and the Bigger Data Management Picture

Data Management

- Metadata Management is directly tied to the larger Data Management topic
- **Data Management:** The systematic processes and governance procedures for applications, processes, data, and technology at a holistic enterprise perspective
- The purpose of data management is to bring enterprise order, purpose, structure, efficiency, and performance to applications, processes, data, metadata and technology
- Data management is not a single technology or component, but a coordinated framework of disciplines for managing data, metadata and information assets throughout the organization
- ***Data Does Not Manage Itself!!***

We're a State Government…Why Should We Care About Metadata Management?

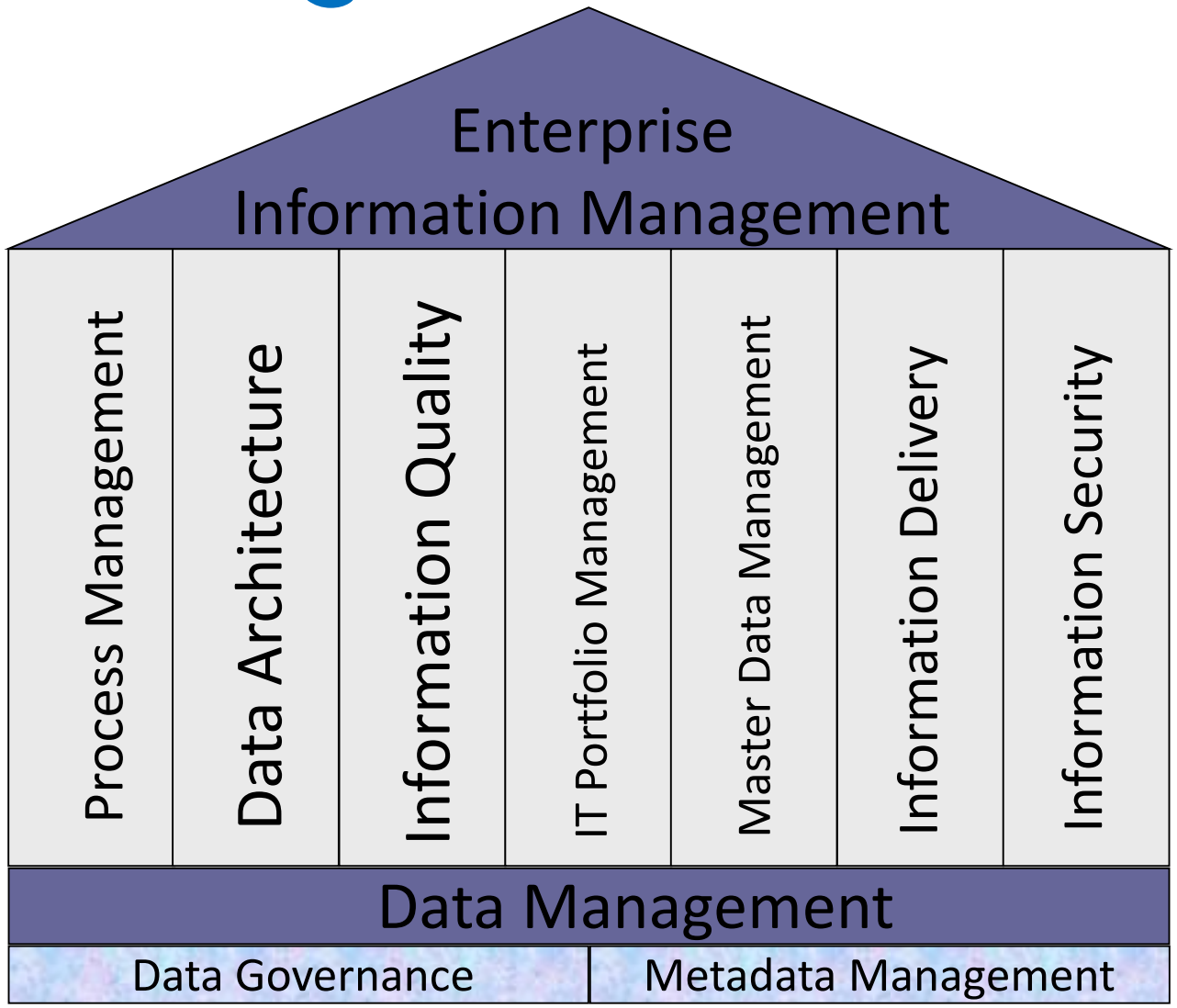
- Agile software development
  - Hard to be agile when you don't know the Who, What, When, Where, How and Why of your data
- Multisourcing
  - This is a significant challenge. At the Federal level some contracts go through 4 – 5 different vendors until it reaches the one doing the work
  - Making sure all of these vendors have access to our metadata is vital to ensure that the various project components actually work together

We're a State Government…Why Should We Care About Metadata Management?

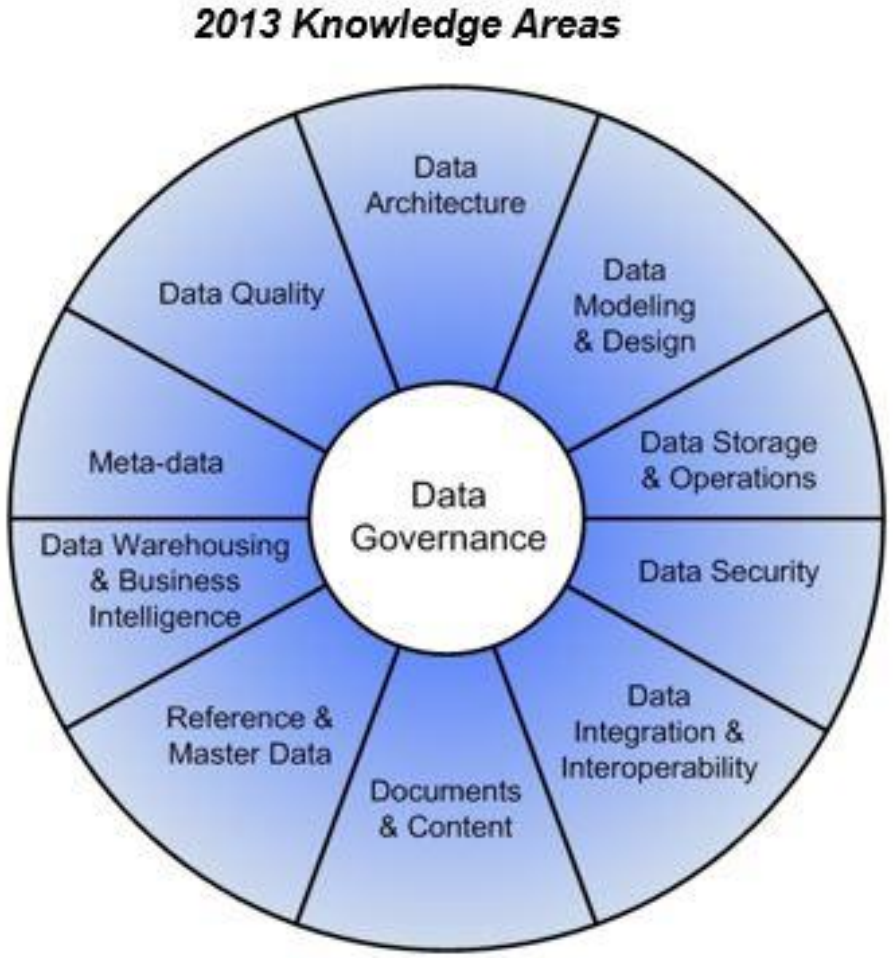
- Cross jurisdictional collaboration
  - If we have proper metadata management we can easily share communicate our data and its meaning to our fellow state governments
  - If we don't have this capability all of this communication will need to be done manually with meetings and phone calls
  - As a result, it won't get done
- Federal requirements and oversight continue to encroach and place demands on state governments
  - We need to have our data properly managed so that we can meet these ever changing requirements
  - There will be penalties associated with non-compliance

Knowledge Areas

Data Management is the foundation for all of the other Enterprise Information Management focus areas. Regardless of which focus area you target first, you will need to do Data Management



DMBOK Knowledge Areas



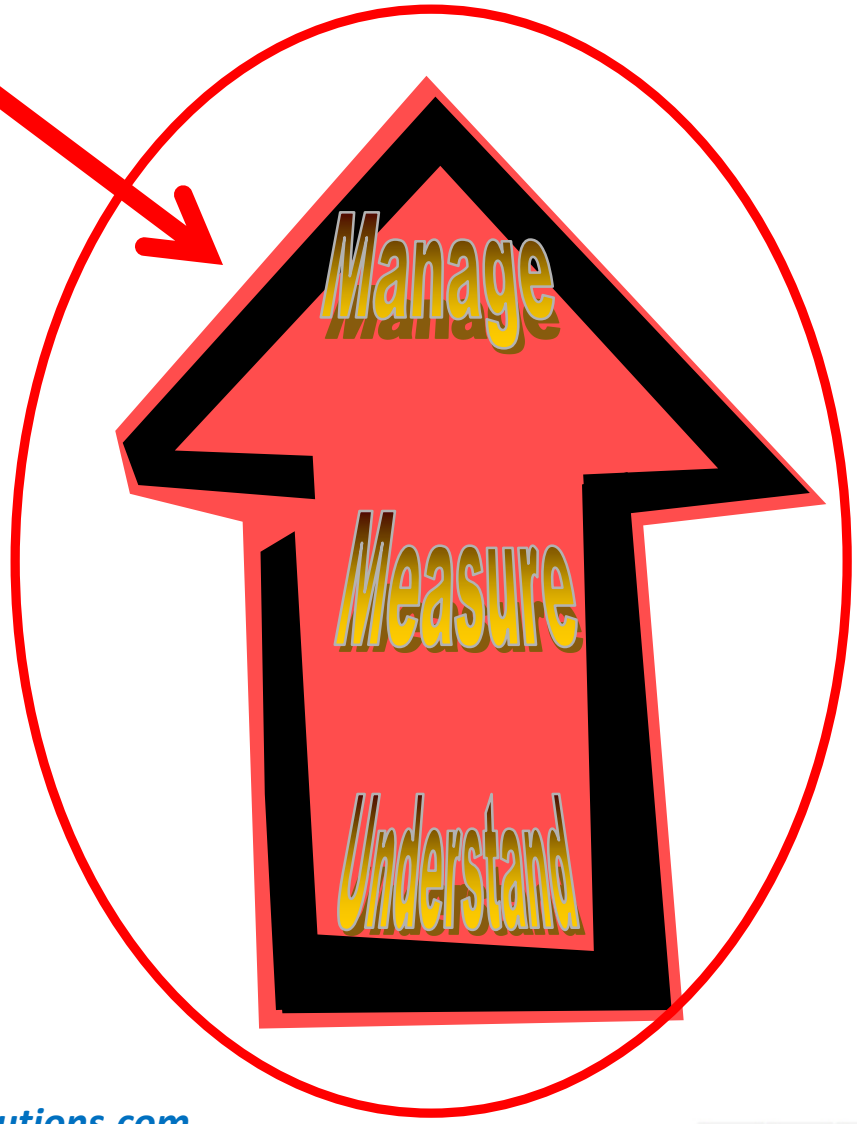
How Do You Manage Information Assets?

This is all Metadata Management

You cannot manage what you do not measure

You cannot measure what you do not understand

You do not understand…….

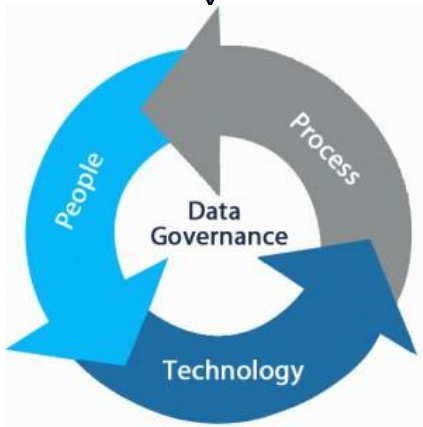


Metadata Management Environment

Metadata Management Environment

Data Security & Compliance

Machine Learning & AI



Advanced Analytics

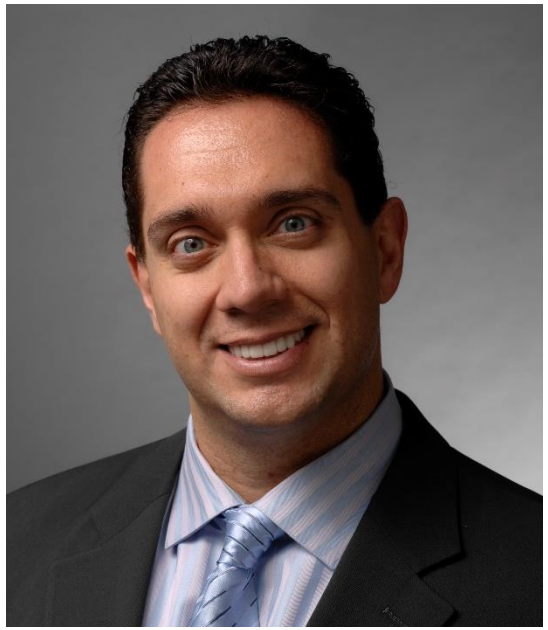
Questions

Contact Information

David Marco, CDMP, CDP, CBIP, MBA

DMarco@ewsolutions.com

http://www.EWSolutions.com



@EWSolutions /EWSolutions EWSolutions EWSolutions

GOVERNANCE METADATA ANALYTICS EWSolutions METHODOLOGIES EWSOLUTIONS.COM