Repository Redux: Lessons Learned from 40 Years of Metadata Repositories

ABSTRACT

The data dictionary/metadata repository concept—originally delivered in the 1960s as a specialized bill-of-materials database to keep track of the explosion of interrelated and functionally redundant components and relationships in already complex masterfiles and databases—fell into disuse with the rise of client-server computing in the late 1980s.

With easily accessible documentation—what are the validation rules for this field?—in properly implemented and automatically updated metadata repository(s), systems become more flexible and responsive to constantly changing market demands.

The business and technical challenge remains the same: when something needs to be changed, what interconnected components are impacted? Since change is constant, plan and prepare for it. Do not regard it as an unexpected surprise.

Purpose of this presentation is to expose those unfamiliar with the challenges of the metadata repository genre to the potentially unanticipated requirements for success—content, access, automated change capture, and rationalization/standards.

BIOGRAPHY

David Eddy
General Manager
David Eddy & Associates

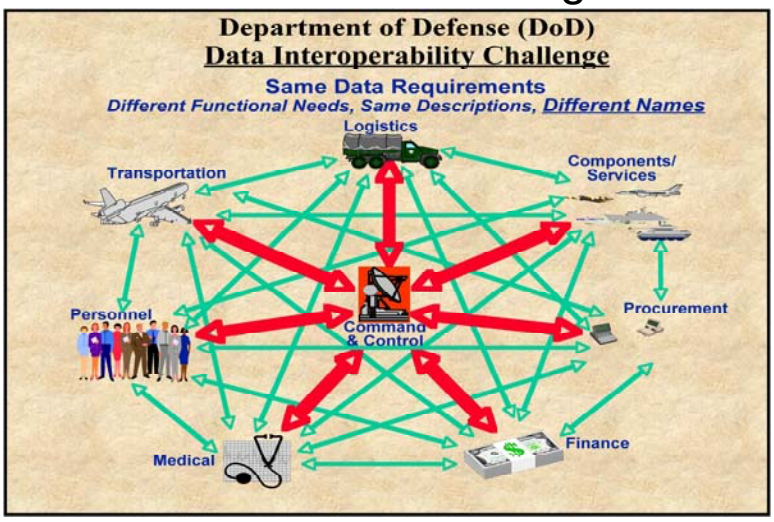
David Eddy is a consult specializing in missionary marketing and sales efforts for niche information resource management products. Mr. Eddy has 35+ years in commercial software development and software sales & marketing. In 1988 he entered the rapidly collapsing metadata repository market. He enabled one client to expand to Europe and Japan.

The collapse of the repository tools market in 1993 led Mr. Eddy to the Y2K inventory and impact analysis market, where he was an early thought leader. He published 110 weekly Y2K articles, was interviewed by the Wall Street Journal in August 1995, briefed Senator Daniel Patrick Moynihan in 1996, and testified to a Senate committee in 1998.

Mr. Eddy has an M.B.A. from Babson College and a B.A. in Russian history from Union College.

Repository Redux:
A Brief Walk Though History

A Common Challenge



• Source: Whitemarsh Information Systems (www.wiscorp.com)

July 14, 2011 MIT- IQIS 2

Rise of the DBMS

- Products & Vendors:
  - DatacomDB (Applied Data Research) 1959
  - IDMS/IDD (Cullinet) 1960s
  - IMS IBM + Rockwell + Caterpillar for Apollo Moon shot 1966
    - ➔ 2008 - Generates US$700M in annual revenues ←
  - Cincom (Cincom) 1968
  - ADABAS (Software AG) 1970
  - Model 204 (Computer Corporation of America) 1972
  - Oracle (Oracle) 1977
- Dictionaries - Active vs Passive.
- You're a single DBMS shop, right?

July 14, 2011 MIT- IQIS 3

Data Dictionary => Metadata Repository

- Formal definition of repository: "burial vault."
- "Data Dictionary" was the industry term 1965 - 1989.
- IBM's AD/Cycle gave us "Metadata Repository" ~1990.

July 14, 2011 MIT- IQIS 4

The Data Model

It's JUST Documentation

ENTERPRISE contains SYSTEMS

SYSTEM contains PROGRAMS

PROGRAM contains DATA STRUCTURES

LOGIC

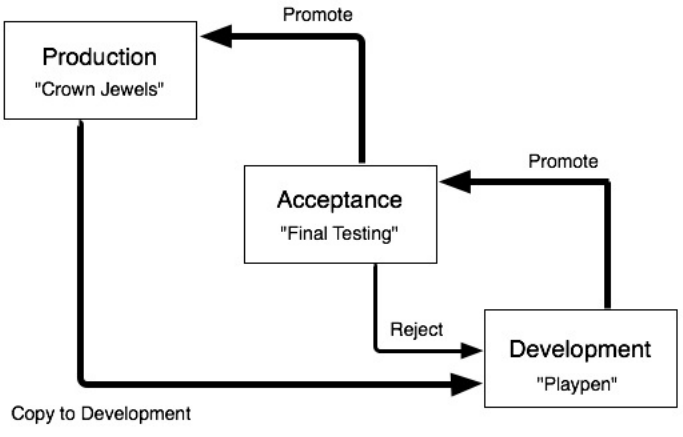
LINKS [to other PROGRAMS]

DATA STRUCTURE contains DATA ELEMENTS

DATA ELEMENT contains RULES.

July 14, 2011 MIT- IQIS 5

Software Configuration Management

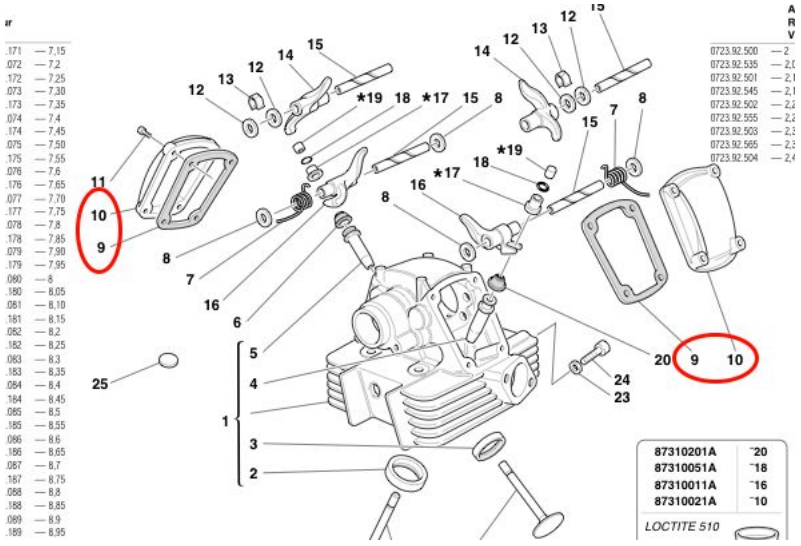


July 14, 2011 MIT- IQIS 6

How Many Moving Pieces?

	May '04	Dec '04	Nov '05	Sep '06	Mar '09	5 Year Change
DB2 Plan	2,834	3,176	3,493	4,027	4,778	68.6%
TSO Parm Card (SYSTSIN)	5,983	6,670	7,154	8,170	9,425	57.5%
Copybook	22,761	25,944	27,139	28,987	31,327	37.6%
DB2 Column	42,297	50,444	51,859	53,225	57,325	35.5%
Program	30,339	32,469	33,478	35,085	40,950	35.0%
Parameter Card	28,716	31,541	33,779	37,755	38,171	32.9%
DB2 Table	22,660	28,001	28,302	28,702	29,778	31.4%
Dataset	87,578	100,911	104,973	119,307	106,343	21.4%
Load Module	36,148	38,408	39,640	41,704	43,362	20.0%
DB2 Database	178	184	191	201	211	18.5%
Job Name	24,965	28,007	30,096	34,619	28,550	14.4%
Procedure	19,185	20,704	22,036	24,083	21,899	14.1%
Job Step	124,617	134,405	141,834	163,162	137,730	10.5%
Job	29,804	33,065	34,325	39,148	31,639	6.2%
Application	773	782	874	878	813	5.2%
DDNAME	888,956	953,998	991,545	1,115,416	898,692	1.1%
IMS Segment	692	678	670	671	673	-2.7%
IMS PSB	12,870	11,430	11,503	11,825	12,226	-5.0%
IMS PSBname	8,230	6,778	6,895	7,164	7,454	-9.4%
IMS Database	527	481	476	476	449	-14.8%
	~~ 43 rows deleted ~~~					
	1,481,884	1,604,650	1,668,192	1,855,160	1,599,121	7.9%

Bill-of-Materials Parts Explosion



Impact Analysis is Forever!

- Bill-of-Materials thinking is a challenge for paper intensive service firms, particularly when the parts are invisible bits of software.
- Difficult to find what's already there… easier to make another one, and another one, & another one…
- The numbers of components and interrelationships are HUGE!
- Modern software style is more, smaller fragments.
- Why were Y2K inventories done from scratch?
- What happened to the Y2K inventories?

- Ability to answer impact questions is Repository value.

Database Structure

FName	LName	DoB	Address
A	Lincoln	02-12-1809	a.lincoln@whitehouse.gov
T	Roosevelt	10-27-1858	t_roosevelt@whitehouse.gov
F	Roosevelt	01-30-1882	FDRoosevelt@whitehouse.gov
W	Clinton	08-19-1946	wjclinton@whitehouse.gov

Metadata Repository Structure

Address	<pointer>	<pointer>	<pointer>…
DoB	<pointer>	<pointer>	<pointer>…
FName	<pointer>	<pointer>	<pointer>…
LName	<pointer>	<pointer>	<pointer>…

Semantics - the Hard Part

- How many names/synonyms/labels/tags?
  - Life insurance company with 70 different names for "policy number"… in 1980.
  - You can start/begin/commence any where you want.
  - Acronyms dictionary: 2,000 terms, 68,000 meanings.
  - An organization has between 1,500 to 6,000 concepts.
  - EDS found Navy with 100,000 "systems"… redundancy?
- Making sense of this Gordian Knot is NOT impossible.

- The idea of "always milk in the milk jug" is a foreign concept to most of the systems we've built to date.

Extra Credit Reading

- Long term success story. CSF - Hire a Marine. (www.tdan.com/view-articles/6123)
- Whitemarsh Systems (www.wiscorp.com)
- "Ambiguous Words" (www.kurzweilai.net/ambiguous-words)
- VisualThesaurus (www.visualthesaurus.com)

Thank you.

?

David Eddy
deddy@davideddy.com
781-455-0949