



# ***DATA RESOURCE QUALITY***

## ***TURNING BAD HABITS INTO GOOD PRACTICES***

***NATIONAL FORUM ON EDUCATION STATISTICS  
FEBRUARY 21, 2005  
NEW ORLEANS***

***MICHAEL H. BRACKETT***

***CONSULTING DATA ARCHITECT  
DATA RESOURCE DESIGN & REMODELING***



P.O. Box 151  
Lilliwaup, Washington 98555  
Voice 360 877 5507  
Fax None  
mhbrackett@aol.com  
members.aol.com/mhbrackett



# ***CONTENTS***

- ◆ ***CURRENT SITUATION***
- ◆ ***DATA RESOURCE CONCEPTS***
- ◆ ***HALTING DATA DISPARITY***
- ◆ ***RESOLVING DATA DISPARITY***
- ◆ ***CULTURAL CONSIDERATIONS***
- ◆ ***CONCLUSION***
- ◆ ***PRESENTER'S BIBLIOGRAPHY***



# ***CURRENT SITUATION***

## ◆ ***BUSINESS INFORMATION DEMAND***

***AN ORGANIZATION'S CONTINUOUSLY INCREASING, CONSTANTLY CHANGING NEED FOR CURRENT, ACCURATE, INTEGRATED INFORMATION, OFTEN ON SHORT NOTICE OR VERY SHORT NOTICE, TO SUPPORT ITS BUSINESS ACTIVITIES***

## ◆ ***CURRENT BUSINESS IMPACT***

- **THIS BUSINESS INFORMATION DEMAND IS NOT BEING MET**
- **PUBLIC & PRIVATE SECTOR BUSINESS IS IMPACTED**
- **BUSINESS DECISIONS ARE LESS-INFORMED**
- **CITIZENS & CUSTOMERS ARE IMPACTED**
- **OPPORTUNITIES ARE MISSED**

## ◆ ***PROBLEM***

- **HIGH QUANTITIES OF DISPARATE DATA IN MOST ORGANIZATIONS**
- **DATA NOT THOROUGHLY UNDERSTOOD**
- **DATA RESOURCE NOT FORMALLY MANAGED**
- **DATA NOT READILY AVAILABLE**
- **DATA QUALITY LOW**

***BUSINESS INFORMATION DEMAND NOT BEING MET!***



## ***CURRENT SITUATION***

### ◆ ***DISPARATE DATA - A TRUISM***

***DATA THAT ARE ESSENTIALLY NOT ALIKE, OR ARE DISTINCTLY DIFFERENT IN KIND, QUALITY, OR CHARACTER. THEY ARE UNEQUAL AND CANNOT BE READILY INTEGRATED TO ADEQUATELY MEET THE BUSINESS INFORMATION DEMAND.***

### ◆ ***BASIC DATA PROBLEM***

- **UNKNOWN DATA EXISTANCE**
  - ORGANIZATION NOT AWARE OF ALL DATA AT ITS DISPOSAL
  - USUALLY NOT EVEN INVENTORIED
  
- **UNKNOWN DATA MEANING**
  - CONTENT AND MEANING OF DATA NOT FULLY KNOWN
  - DATA NOT THOROUGHLY UNDERSTOOD
  
- **HIGH DATA REDUNDANCY**
  - DATA HIGHLY REDUNDANT AND INCONSISTENT
  - AVERAGE REDUNDANCY FACTOR OF 10 FOR LARGE ORGANIZATIONS
  
- **HIGH DATA VARIABILITY**
  - DATA HIGHLY VARIABLE IN FORMAT AND CONTENT
  - AVERAGE FACTOR OF 15 TO 20 FOR LARGE ORGANIZATIONS

***THAT'S THE GOOD NEWS!***



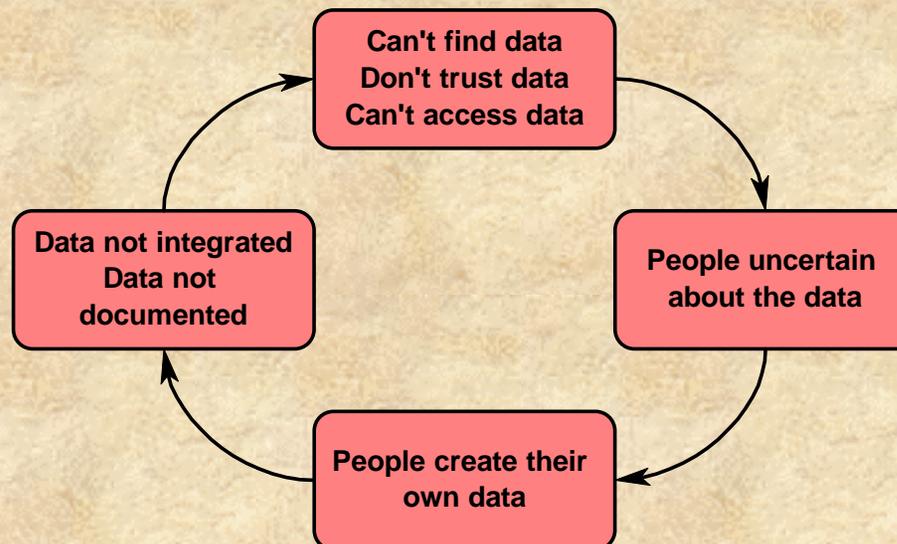
## ***CURRENT SITUATION***

### ◆ ***DISPARATE DATA RESOURCE***

***A DATA RESOURCE THAT IS SUBSTANTIALLY COMPOSED OF DISPARATE DATA THAT ARE DIS-INTEGRATED AND NOT SUBJECT ORIENTED. A STATE OF DISARRAY WHERE THE LOW QUALITY DOES NO, AND CANNOT, ADEQUATELY SUPPORT THE BUSINESS INFORMATION DEMAND.***

### ◆ ***DISPARATE DATA CYCLE***

***A SELF-PERPETUATING CYCLE WHERE DISPARATE DATA CONTINUE TO BE PRODUCED AT AN EVER-INCREASING RATE BECAUSE PEOPLE DO NOT KNOW ABOUT EXISTING DATA OR DO NOT WANT TO USE EXISTING DATA.***



***THAT'S THE BAD NEWS!***



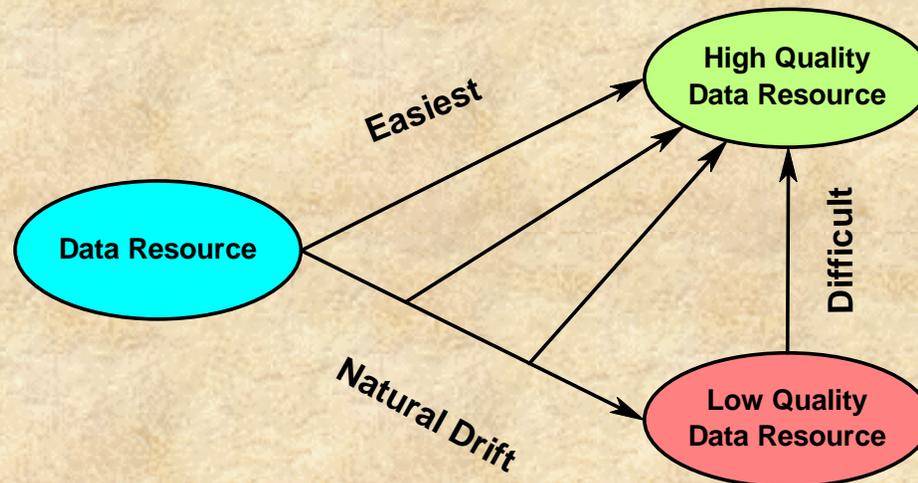
## ***CURRENT SITUATION***

### ◆ ***DISPARATE DATA SPIRAL***

- **DISPARATE DATA INCREASING BY SEVERAL ORDERS OF MAGNITUDE**
- **CREATING DISPARATE DATA FASTER THAN EVER BEFORE**
- **SPREADING TO NON-TABULAR DATA AND DATA RESOURCE DATA**
- **DISPARITY SPIRALING OUT OF CONTROL**
- **NO FORESEEABLE END IN SIGHT!**

### ◆ ***DATA RESOURCE DRIFT***

***THE NATURAL, STEADY DRIFT OF A DATA RESOURCE TOWARDS DISPARITY IF ITS DEVELOPMENT IS NOT PROPERLY MANAGED AND CONTROLLED.***



***THAT'S THE REALLY BAD NEWS!***



## ***CURRENT SITUATION***

### ◆ ***DATA DILEMMA***

- ORGANIZATIONS FACING REAL DILEMMA ABOUT THEIR DATA RESOURCE
- HIGH DEMAND FOR INTEGRATED DATA TO SUPPORT BUSINESS NEEDS
- CONTINUED RAPID PRODUCTION OF DISPARATE DATA

### ◆ ***NO STATUS QUO***

- THERE IS NO STATUS QUO FOR DEVELOPING A HIGH-QUALITY DATA RESOURCE
- NATURAL DRIFT TOWARD DISPARITY WILL CONTINUE
- SITUATION WILL GET WORSE

***A STATUS QUO LEADS TO  
ORGANIZATIONAL FAILURE BY INFORMATION DEPRIVATION!***



## ***CURRENT SITUATION***

### ◆ ***MUST CONSCIOUSLY ALTER THE NATURAL DRIFT***

- TOWARD HIGH-QUALITY DATA RESOURCE
- BREAK DISPARATE DATA CYCLE
- STOP THE SPIRALING DISPARITY
- CREATE A HIGH-QUALITY SHARABLE DATA RESOURCE

### ◆ ***SILVER BULLETS JUST DO NOT EXIST***

OBJECTIVE: ***AN ATTEMPT TO ACHIEVE SOME GAIN WITHOUT ANY PAIN***

RESULT: ***ENDURING CONSIDERABLE PAIN WITH MINIMAL GAIN***

***THERE ARE NO SILVER BULLETS!***



# ***DATA RESOURCE CONCEPTS***

## ◆ ***BUSINESS INTELLIGENCE VALUE CHAIN***



## ◆ ***I-ORGANIZATION***

- **INTELLIGENT LEARNING ORGANIZATION - HUMAN RESOURCE REALM**
- **HUMAN RESOURCE REALM LINKS TECHNOLOGY REALM & BUSINESS REALM**
- **REQUIRES GOOD INFORMATION TO SUPPORT BUSINESS STRATEGIES & GOALS**
- **REQUIRES A HIGH-QUALITY SHARABLE DATA RESOURCE**

***MUST BEGIN WITH THE DATA RESOURCE!***



# ***DATA RESOURCE CONCEPTS***

## **◆ MANY PEOPLE STILL CONFUSE 'DATA' - 'INFORMATION' - 'KNOWLEDGE'**

- **DATA**
  - INDIVIDUAL FACTS OUT OF CONTEXT WITH LITTLE MEANING
  - 123.45
  
- **DATA IN CONTEXT**
  - INDIVIDUAL FACTS WITH MEANING
  - ACCOUNT BALANCE AT NOON ON JANUARY 12, 2001
  
- **INFORMATION**
  - SET OF DATA IN CONTEXT
  - RELEVANT TO ONE OR MORE PEOPLE
  - AT A POINT IN TIME OR FOR A PERIOD OF TIME
  
- **KNOWLEDGE**
  - INFORMATION RETAINED BY INDIVIDUALS
  - COMBINED WITH EXPERIENCE

***DATA ARE THE FOUNDATION!***



# ***DATA RESOURCE CONCEPTS***

## **◆ QUALITY DEFINED**

- **DATA RESOURCE QUALITY**

***A MEASURE OF HOW WELL THE DATA RESOURCE SUPPORTS THE CURRENT AND THE FUTURE BUSINESS INFORMATION DEMAND***

- **DATA QUALITY**

***A SUBSET OF DATA RESOURCE QUALITY DEALING WITH DATA VALUES***

- **ULTIMATE DATA RESOURCE QUALITY**

***A DATA RESOURCE THAT IS STABLE ACROSS CHANGING BUSINESS AND CHANGING TECHNOLOGY SO IT CONTINUES TO SUPPORT THE CURRENT AND FUTURE BUSINESS INFORMATION DEMAND***

- **INFORMATION QUALITY**

***A MEASURE OF THE ABILITY TO GET **THE RIGHT DATA**, TO THE RIGHT PEOPLE, IN THE RIGHT PLACE, AT THE RIGHT TIME, IN THE RIGHT FORM, AT THE RIGHT COST, SO THEY CAN MAKE THE RIGHT DECISIONS, AND TAKE THE RIGHT ACTIONS.***

***INTELLIGENT LEARNING ORGANIZATION  
NEEDS A HIGH-QUALITY STABLE DATA RESOURCE!***



# DATA RESOURCE CONCEPTS

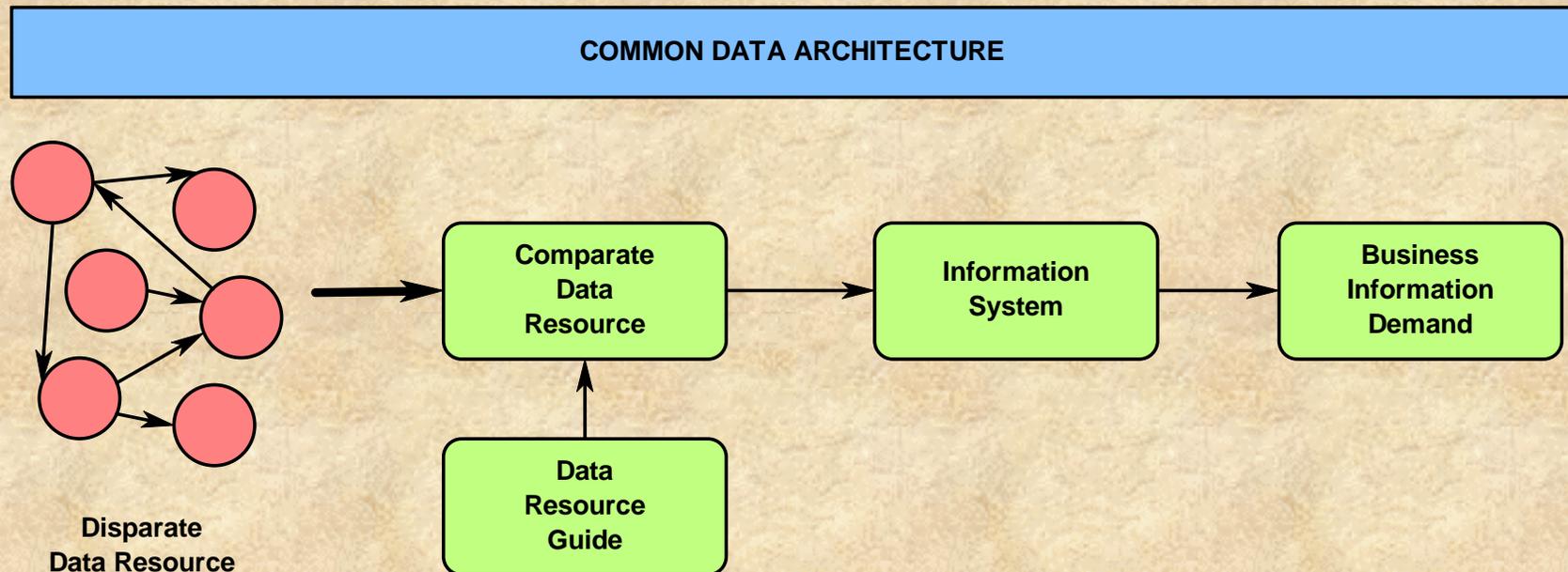
## ◆ COMPARE DATA VISION

- COMPARE DATA

*DATA THAT ARE ALIKE, SIMILAR IN KIND, QUALITY, AND CHARACTER, ARE EASILY UNDERSTOOD, AND CAN BE READILY INTEGRATED*

- COMPARE DATA RESOURCE

*SUBJECT ORIENTED, INTEGRATED, HIGH QUALITY, SHARABLE DATA RESOURCE THAT FULLY SUPPORTS THE CURRENT AND THE FUTURE BUSINESS INFORMATION DEMAND*

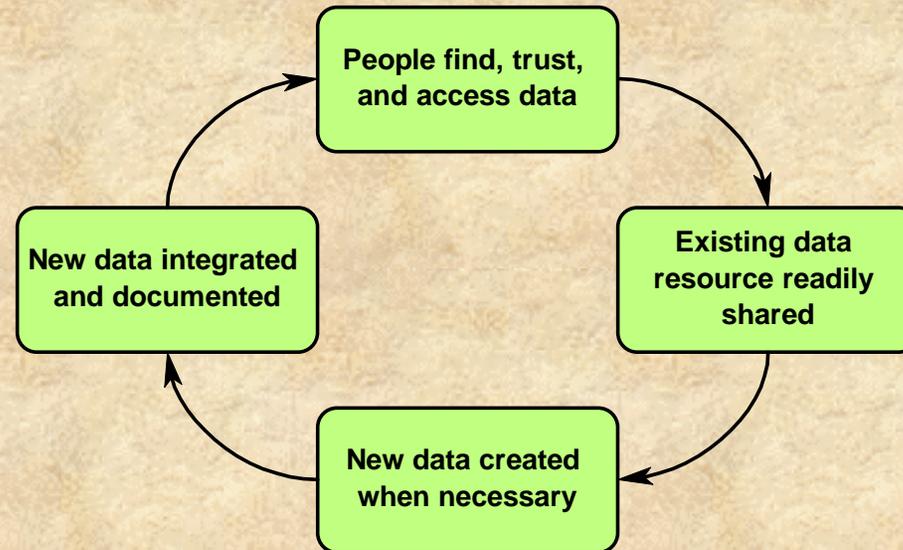




# ***DATA RESOURCE CONCEPTS***

## **◆ COMPARETE DATA CYCLE**

**A SELF-PERPETUATING CYCLE WHERE THE USE OF COMPARTE DATA IS CONTINUALLY REINFORCED BECAUSE PEOPLE UNDERSTAND AND TRUST THE DATA.**



## **◆ RESOLVE THE DATA DILEMMA**

- **BRING STABILITY TO THE DATA RESOURCE TO TRANSCEND CHANGE**

***STOP THE NATURAL DRIFT OF A DATA RESOURCE  
BY HALTING THE DISPARATE DATA CYCLE  
AND STARTING A COMPARTE DATA CYCLE!***



# ***DATA RESOURCE CONCEPTS***

## **◆ COMPARATE DATA RESOURCE**

- **SUBJECT ORIENTED**
  - **BASED ON BUSINESS OBJECTS AND EVENTS IN THE REAL WORLD**
  - **DATA SUBJECTS REPRESENT BUSINESS OBJECTS AND EVENTS**
  - **DATA RESOURCE STRUCTURED BY DATA SUBJECTS**
  - **ALL CHARACTERISTICS ABOUT A DATA SUBJECT STORED WITH THE DATA SUBJECT**
  
- **INTEGRATED**
  - **FULLY INTEGRATED WITHIN A COMMON DATA ARCHITECTURE**
  - **ONE ARCHITECTURE FOR THE ENTIRE DATA RESOURCE**
  - **ALL DATA MANAGED WITHIN THAT SINGLE ARCHITECTURE**
  - **DATA PROPERLY DEPLOYED TO MEET BUSINESS NEEDS**
  
- **SUPPORTED BY DATA RESOURCE GUIDE**
  - **COMPREHENSIVE INDEX TO DATA RESOURCE**
  - **A COMPLETE INTEGRATED UNDERSTANDING ABOUT THE DATA RESOURCE**
  
- **INTEGRATED**
  - **ACROSS ORGANIZATION UNITS**
  - **ACROSS BUSINESS ACTIVITIES**

***AN INTEGRATED DATA RESOURCE  
IS ONE VERSION OF TRUTH ABOUT THE BUSINESS!***



# ***DATA RESOURCE CONCEPTS***

## **◆ DATA ARCHITECTURE DEFINITION 1**

**THE METHOD OF DESIGN AND CONSTRUCTION OF A DATA RESOURCE THAT IS BUSINESS DRIVEN, BASED ON REAL-WORLD SUBJECTS PERCEIVED BY THE ENTERPRISE, AND IMPLEMENTED INTO APPROPRIATE OPERATING ENVIRONMENTS. IT CONSISTS OF COMPONENTS THAT PROVIDE A CONSISTENT FOUNDATION ACROSS ORGANIZATIONAL BOUNDARIES TO PROVIDE EASILY IDENTIFIABLE, READILY AVAILABLE HIGH-QUALITY DATA TO SUPPORT THE BUSINESS INFORMATION DEMAND**

## **◆ DATA ARCHITECTURE DEFINITION 2**

**THE COMPONENT OF THE DATA RESOURCE FRAMEWORK THAT CONTAINS ALL THE ACTIVITIES, AND THE PRODUCTS OF THOSE ACTIVITIES, RELATED TO THE IDENTIFICATION, NAMING, DEFINITION, STRUCTURING, INTEGRITY ACCURACY, EFFECTIVENESS, AND DOCUMENTATION OF THE DATA RESOURCE**

**DATA INTEGRATION REQUIRES A FORMAL DEFINITION  
OF DATA ARCHITECTURE!**



# ***DATA RESOURCE CONCEPTS***

## **◆ COMMON DATA ARCHITECTURE**

### **● COMMON CONTEXT FOR**

- INVENTORYING ALL DATA
- UNDERSTANDING THE CONTENT AND MEANING OF DATA
- IMPROVING DATA QUALITY
- TRANSFORMING AND INTEGRATING DISPARATE DATA
- DEFINING NEW DATA
- MANAGING DYNAMIC DATA DEPLOYMENT

### **● EINSTEIN'S PRINCIPLE**

- A PROBLEM CANNOT BE SOLVED WITH THE SAME LEVEL OF TECHNOLOGY USED TO CREATE THAT PROBLEM
- A HIGHER LEVEL OF TECHNOLOGY IS NEEDED
- THE COMMON DATA ARCHITECTURE IS THE HIGHER LEVEL OF TECHNOLOGY FOR UNDERSTANDING AND MANAGING DATA

### **● DEVELOP CUSTOMIZED METHOD**

- ONE SIZE FITS ALL METHODS DO NOT WORK
- COMMON DATA ARCHITECTURE HAS CONCEPTS - PRINCIPLES - TECHNIQUES
- ORGANIZATION HAS PROBLEMS - NEEDS - ENVIRONMENT
- CUSTOMIZE A METHOD TO SUIT ORGANIZATION

***CUSTOMIZED METHOD HELPS RESOLVE THE DATA DILEMMA!***



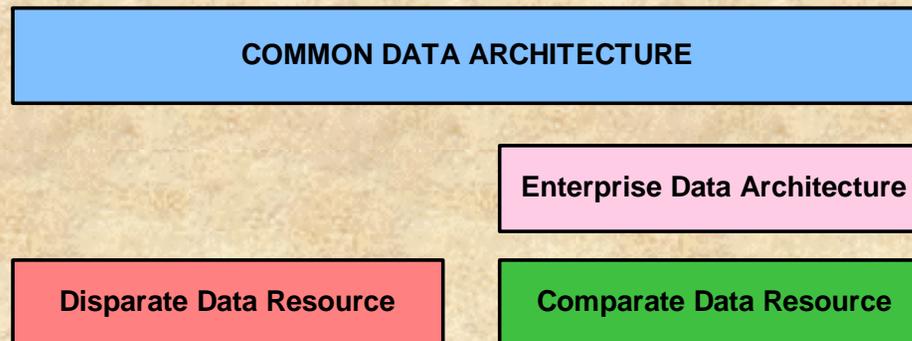
# ***DATA RESOURCE CONCEPTS***

## **◆ ENTERPRISE DATA ARCHITECTURE**

- **MANY DIFFERENT DEFINITIONS**
- **CONFLICTING AND CONFUSING**

***AN ENTERPRISE DATA ARCHITECTURE IS A SUBSET OF THE COMMON DATA ARCHITECTURE THAT REPRESENTS THE COMPARATE DATA RESOURCE.***

- **IT REPRESENTS THE DESIRED ENTERPRISE-WIDE DATA ARCHITECTURE FOR AN ORGANIZATION.**
- **IT IS THE PATTERN FOR BUILDING THE COMPARATE DATA RESOURCE FROM THE DISPARATE DATA**



***THE ENTERPRISE DATA ARCHITECTURE REPERESENTS  
THE COMPARATE DATA RESOURCE!***



# ***DATA RESOURCE CONCEPTS***

## **◆ *ACHIEVING AN INTEGRATED DATA RESOURCE***

- **PHASE 1 - HALT THE CREATION OF DISPARATE DATA**
  - ESTABLISH A COMMON DATA ARCHITECTURE
  - DEVELOP NEW DATA WITHIN THAT COMMON DATA ARCHITECTURE
  - PROACTIVE PHASE
  - PREVENTS FURTHER DATA DISPARITY
  - MUST BE DONE FIRST
  
- **PHASE 2 - RESOLVE THE EXISTING DATA DISPARITY**
  - TRANSITION DISPARATE DATA TO A COMPARETE DATA RESOURCE
  - FORMAL DATA TRANSFORMATION WITHIN A COMMON DATA ARCHITECTURE
  - REACTIVE PHASE
  - RESOLVES DISPARATE DATA
  - MUST BE DONE AFTER PHASE 1

***DATA INTEGRATION BEGINS BY  
STOPPING FURTHER DATA DISPARITY  
AND THEN RESOLVING EXISTING DISPARITY!***



# ***DATA NAMES***

## ◆ **FORMAL DATA NAMES**

***DATA NAMES THAT READILY AND UNIQUELY IDENTIFY A FACT OR GROUP OF FACTS IN THE DATA RESOURCE. THEY ARE DEVELOPED WITHIN A FORMAL DATA NAMING TAXONOMY AND ARE ABBREVIATED WHEN NECESSARY WITH A FORMAL SET OF ABBREVIATIONS AND A FORMAL ABBREVIATION ALGORITHM.***

## ◆ **GOOD PRACTICES**

### **FORMAL DATA NAMING TAXONOMY**

- **MAKES DATA NAMING EQUIVALENT TO ANIMALS, PLANTS, MINERALS, CHEMICALS**

### **DATA NAMING VOCABULARY**

- **EXPAND CLASS WORD CONCEPT TO ALL COMPONENTS OF THE DATA NAMING TAXONOMY**
- **SET OF COMMON WORDS FOR EACH COMPONENT OF DATA NAMING TAXONOMY**
- **PROVIDE COMMON MEANING ACROSS DATA NAMES**

### **PRIMARY DATA NAME**

***THE FORMAL DATA NAME THAT IS FULLY SPELLED OUT, REAL WORLD, UNABBREVIATED, UNTRUNCATED, BUSINESS NAME OF THE DATA THAT HAS NO SPECIAL CHARACTERS OR LENGTH LIMITATIONS.***

- **MUST BE UNIQUE AND MEANINGFUL TO THE BUSINESS**
- **ALL OTHER DATA NAMES ARE ALIASES OF THE PRIMARY DATA NAME**



# ***DATA NAMES***

## **DATA NAME WORD ABBREVIATIONS**

- PRIMARY DATA NAMES MUST BE FORMALLY ABBREVIATED TO MEET LENGTH RESTRICTIONS
- DATA NAME WORD ABBREVIATIONS

## **DATA NAME ABBREVIATION ALGORITHM**

- A FORMAL PROCEDURE FOR ABBREVIATING PRIMARY DATA NAMES USING ABBREVIATIONS
- VARIETY OF DIFFERENT ALGORITHMS
- CAN HAVE MORE THAN ONE ABBREVIATION ALGORITHM
- THE NUMBER OF ALGORITHMS IS NOT IMPORTANT - THE FORMALITY IS



# ***DATA DEFINITIONS***

## ◆ **COMPREHENSIVE DATA DEFINITIONS**

**FORMAL DATA DEFINITIONS THAT PROVIDE A COMPLETE, MEANINGFUL, EASILY READ, READILY UNDERSTOOD DEFINITION THAT THOROUGHLY EXPLAINS THE CONTENT AND MEANING OF THE DATA.**

## ◆ **GOOD PRACTICES**

### **MEANINGFUL DATA DEFINITIONS**

- **CONTENT AND MEANING WITH RESPECT TO THE BUSINESS**
- **NOT WHERE DATA STORED, HOW DATA USED, OR WHERE DATA CAPTURED**
- **SIMILARITIES AND DIFFERENCES WITH CLOSELY RELATED DATA**
- **HISTORY IF PERTINENT TO THE MEANING - NOT AN AUDIT TRAIL OF DEFINITIONS**
- **WHAT IS AND IS NOT INCLUDED**
- **INCLUDE BUSINESS CLIENTS**

### **THOROUGH DATA DEFINITIONS**

- **NO LENGTH LIMITATIONS**
- **APPROPRIATE BUSINESS TERMINOLOGY**
- **STRONG DENOTATIVE MEANING - LIMIT CONNOTATIVE MEANING**
- **1 TO 3 PARAGRAPHS OF 2 TO 3 SENTENCES EACH**

### **CORRECT DATA DEFINITIONS**

- **CURRENT WITH THE BUSINESS**
- **ROUTINELY REVIEWED AND ENHANCED**
- **EVOLVE AS THE BUSINESS EVOLVES**



# ***DATA STRUCTURES***

## ◆ **PROPER DATA STRUCTURES**

***A DATA STRUCTURE THAT PROVIDES A SUITABLE REPRESENTATION OF THE BUSINESS, AND DATA SUPPORTING THE BUSINESS, THAT IS RELEVANT TO THE INTENDED AUDIENCE.***

## ◆ **GOOD PRACTICES**

### **DATA STRUCTURE COMPONENTS**

- **ENTITY - RELATION DIAGRAM**  
SHOWS THE DATA SUBJECTS AND THE RELATIONS BETWEEN THEM  
LIKE THE FLOOR PLAN OF A HOUSE  
DOES NOT SHOW DETAIL WITHIN EACH DATA SUBJECT  
SYMBOLS FOLLOW SEMIOTIC THEORY
  
- **DATA ATTRIBUTE STRUCTURE**  
SHOWS DATA ATTRIBUTES IN A DATA SUBJECT  
SHOWS ROLES DATA ATTRIBUTES PLAY



# ***DATA STRUCTURES***

## **◆ PRINCIPLES**

- **TECHNICALLY CORRECT & CULTURALLY ACCEPTABLE**
  - **TECHNICALLY CORRECT IS LIKE BUILDING CODES**
  - **CULTURALLY ACCEPTABLE IS LIKE AESTHETICS**
  
- **STRUCTURALLY STABLE & BUSINESS FLEXIBLE**
  - **ARCHITECTURE REMAINS STABLE AS BUSINESS CHANGES**
  - **SUPPORTS BUSINESS PROCESS IMPROVEMENT**
  - **ENCOURAGES BUSINESS IMPROVEMENT**

***TURNING POINT BETWEEN PREDOMINANTLY DISPARATE AND  
PREDOMINANTLY COMPARETE DATA RESOURCE!***



# ***DATA INTEGRITY RULES***

## ◆ ***PRECISE DATA INTEGRITY RULES***

***A DATA INTEGRITY RULE THAT PRECISELY SPECIFIES THE CRITERIA FOR HIGH-QUALITY DATA VALUES AND REDUCES OR ELIMINATES DATA ERRORS.***

## ◆ ***GOOD PRACTICES***

### ***DATA RULE CONCEPT***

- **SUBSET OF BUSINESS RULES ACCORDING TO THE ZACHMAN FRAMEWORK**
- **DEALS WITH THE DATA COLUMN OF THE FRAMEWORK**
- **ALLOWS PRECISE SPECIFICATION FOR QUALITY IMPROVEMENT**
  
- **A DATA DEFINITION IS NOT A DATA RULE**  
**A DEFINITION IS, BY DEFINITION, A DEFINITION**
  
- **PROBLEM MAKING DEFINITION A RULE**  
**IMPACTS A COMPREHENSIVE DEFINITIONS**  
**LEADS TO ONE- LINE PHRASES WITH ADDITIONAL EXPLANATION**  
**ADDITIONAL EXPLANATION OFTEN NOT PROVIDED**  
**REVERSION TO LEGACY DATA DEFINITIONS**

***A DATA DEFINITION IS NOT A DATA RULE!***



# ***DATA INTEGRITY RULES***

## **DATA INTEGRITY RULE NAMES**

- FORMAL NAME ACCORDING TO DATA NAMING TAXONOMY & VOCABULARY
- USES THE ! NOTATION **CUSTOMER. NAME CHANGE!**
- REQUIRES NORMALIZATION OF DATA RULES
  
- DATA RULE IS NORMALIZED  
TO THE DATA RESOURCE COMPONENT IT REPRESENTS  
OR ON WHICH IT TAKES ACTION

## **DATA INTEGRITY RULE NOTATION**

- ACCEPTABLE AND UNDERSTANDABLE TO BUSINESS CLIENTS
- BASED ON MATHEMATICAL AND LOGIC NOTATION WHERE PRACTICAL
- NOTATION READILY AVAILABLE ON STANDARD KEYBOARD

## **DATA INTEGRITY RULE TYPES**

- DATA VALUE
- CONDITIONAL DATA VALUE
- DATA STRUCTURE
- CONDITIONAL DATA STRUCTURE
- DATE DERIVATION
- DATE RETENTION



# ***DATA INTEGRITY RULES***

## **DATA INTEGRITY RULE ENFORCEMENT**

- ONE SET OF DATA INTEGRITY RULES
- APPLY TO ALL DATA ENTERING DATA RESOURCE
- APPLY AS CLOSE TO CAPTURE AS POSSIBLE
- VIOLATIONS ACTIONS DEFINED
- VIOLATION NOTIFICATION NOTIFICATION DEFINED
- DEFAULT VALUES SPECIFIED

## **PROACTIVE DATA QUALITY MANAGEMENT - SELF CORRECTING PROCESS**

- RAPID DATA ERROR IDENTIFICATION
- APPLY TO ENTIRE DATA RESOURCE
- DEFAULT VALUES DOCUMENTED WHEN INSERTED



# ***DATA DOCUMENTATION***

## ◆ ***ROBUST DATA DOCUMENTATION***

***DOCUMENTATION ABOUT THE DATA RESOURCE THAT IS COMPLETE, CORRECT, CURRENT, UNDERSTANDABLE, NON-REDUNDANT, READILY AVAILABLE, AND KNOWN TO EXIST.***

## ◆ ***GOOD PRACTICES***

### ***DATA RESOURCE DATA CONCEPT***

- ***THE TERM **META-DATA** MISUSED AND ABUSED TO THE POINT THAT ITS MEANING IS UNCLEAR***
  
- ***META-DATA FIASCO UNFOLDING***
- ***META-META-DATA, META-META-META-DATA, ETC.***
  
- *****DATA RESOURCE DATA** IS A MUCH BETTER TERM***
- ***ANY DATA NECESSARY FOR THOROUGHLY UNDERSTANDING, FORMALLY MANAGING, AND FULLY UTILIZING THE DATA RESOURCE TO SUPPORT THE BUSINESS INFORMATION DEMAND***
  
- ***DESIGNED AND MANAGED THE SAME AS ANY OTHER DATA - FINANCIAL DATA, CUSTOMER DATA, JUSTICE DATA, HUMAN RESOURCE DATA***
  
- ***SEGMENT OF THE ORGANIZATION'S DATA RESOURCE THE SAME AS ANY OTHER DATA***
  
- ***PART OF THE COMMON DATA ARCHITECTURE***

***DATA RESOURCE DATA DOCUMENT THE ENTIRE DATA RESOURCE!***



# ***DATA DOCUMENTATION***

## **DATA RESOURCE DATA ASPECTS**

- **SEMANTIC DATA RESOURCE DATA**  
BUSINESS ASPECT  
THE MEANING OF THE DATA  
HELP PEOPLE UNDERSTAND AND USE THE DATA RESOURCE  
RELATIVELY DIFFICULT TO CAPTURE AND MAINTAIN
  
- **TECHNICAL DATA RESOURCE DATA**  
PHYSICAL ASPECT  
PARAMETERS ABOUT THE DATA STORAGE AND DATA MANIPULATION  
DATA ABOUT BUILDING, MANAGING, AND MAINTAINING DATABASES  
RELATIVELY EASY TO CAPTURE AND MAINTAIN

***DATA RESOURCE DATA MUST INCLUDE BOTH  
SEMANTIC DATA AND TECHNICAL DATA!***



# ***DATA DOCUMENTATION***

## **COMPLETE DATA DOCUMENTATION**

- INCLUDE SEMANTIC AND TECHNICAL DATA RESOURCE DATA
- INCLUDE ENTIRE SCOPE OF THE DATA RESOURCE
- INCLUDE ALL DATA RESOURCE COMPONENTS

## **CURRENT DATA DOCUMENTATION**

- DOCUMENTATION KEPT CURRENT WITH THE BUSINESS
- ADEQUATELY REPRESENT HISTORICAL CHANGES IN THE BUSINESS / DATA RESOURCE

## **UNDERSTANDABLE DATA DOCUMENTATION**

- UNDERSTANDABLE TO ALL AUDIENCES
- CORRECT DETAIL FOR INTENDED AUDIENCE
- APPROPRIATE PRESENTATION

## **NON-REDUNDANT DATA DOCUMENTATION**

- ONE SINGLE VERSION OF THE TRUTH FOR THE DATA RESOURCE
- DEVELOP WITHIN A COMMON DATA ARCHITECTURE
- SAME CONCEPTS, PRINCIPLES, TECHNIQUES FOR DEVELOPING BUSINESS DATA

## **READILY AVAILABLE**

- READILY AVAILABLE TO ALL AUDIENCES - BUSINESS & TECHNICAL
- EASILY ACCESSIBLE THROUGH BUSINESS APPLICATIONS



# ***DATA DOCUMENTATION***

## **KNOWN TO EXIST**

- ALL AUDIENCES MUST KNOW DATA DOCUMENTATION EXISTS
- CONSTANTLY MAKE PEOPLE AWARE THAT DATA DOCUMENTATION EXISTS

## **ANCILLARY DATA DOCUMENTATION**

- DATA SUBJECT THESAURUS - BUSINESS TERMS POINT TO DATA SUBJECT  
KEYWORD IN CONTEXT & KEY TERM OUT OF CONTEXT
- BUSINESS GLOSSARY  
LIST OF TERMS AND ABBREVIATIONS USED IN THE BUSINESS  
BUSINESS TERMS AND ABBREVIATIONS OFTEN AS DISPARATE AS THE DATA  
HELPS PEOPLE UNDERSTAND THE BUSINESS AND FOCUS ON PREFERRED TERMS



# ***DATA ORIENTATION***

## ◆ **REASONABLE DATA ORIENTATION**

***A DATA ORIENTATION THAT IS PRIMARILY TOWARD THE BUSINESS AND SUPPORT OF BOTH THE CURRENT AND FUTURE BUSINESS INFORMATION DEMAND.***

## ◆ **GOOD PRACTICES**

### **BUSINESS SUBJECT ORIENTATION**

- PRIMARY ORIENTATION TOWARD BUSINESS SUBJECTS
- BUSINESS OBJECTS AND BUSINESS EVENTS IN REAL WORLD

### **BUSINESS CLIENT ORIENTATION**

- DIRECT BUSINESS CLIENT INVOLVEMENT FOR BUSINESS KNOWLEDGE
- CANNOT EXCLUDE BUSINESS KNOWLEDGE
- CANNOT HAVE BUSINESS CLIENTS DRIVE DATABASE DESIGN
- FIND RIGHT MIX OF BUSINESS KNOWLEDGE AND TECHNICAL SKILLS

***COOPERATIVE DEVELOPMENT OF A HIGH-QUALITY DATA RESOURCE!***

### **FIVE-TIER CONCEPT**

- BUILDS ON THE THREE-TIER CONCEPT
- ADDS AN ANALYTICAL TIER AND A PREDICTIVE TIER



# ***DATA ORIENTATION***

## **DATA NORMALIZATION**

- EXPANDED PERSPECTIVE
  - BRINGING DATA INTO NORMAL FORM FOR THE INTENDED PURPOSE
- OPERATIONAL DATA NORMALIZATION  
ANALYTICAL DATA NORMALIZATION  
PREDICTIVE DATA NORMALIZATION

## **SINGLE ARCHITECTURE ORIENTATION**

- ONE INTEGRATED ORGANIZATION-WIDE DATA ARCHITECTURE
- ONE ARCHITECTURE FOR ALL FIVE TIERS AND ALL FIVE SCHEMA
- COMMON DATA ARCHITECTURE CONCEPT

## **SINGLE FACT ORIENTATION**

- SINGLE FACT DATA ATTRIBUTE  
**BIRTH DATE & REASON FOR NO BIRTH DATE**
- SINGLE VALUE DATA ATTRIBUTE  
**SINGLE NAME PER DATA ATTRIBUTE**
- DATA ATTRIBUTE NORMALIZATION  
SINGLE FACT OR GROUP OF CLOSELY RELATED FACTS  
SINGLE VALUE



# ***DATA AVAILABILITY***

## ◆ **ACCEPTABLE DATA AVAILABILITY**

***ENSURE THAT THE DATA ARE AVAILABLE TO MEET THE BUSINESS INFORMATION DEMAND WHILE PROPERLY PROTECTING AND SECURING THOSE DATA.***

## ◆ **GOOD PRACTICES**

### **ADEQUATE DATA ACCESSIBILITY**

- **BALANCE BETWEEN TOO SECURE OR NOT SECURE ENOUGH**
- **PROTECTION FROM UNAUTHORIZED ACCESS, ALTERATION, OR DESTRUCTION**

### **ADEQUATE DATA PROTECTION**

- **REASONABLE PROTECTION AGAINST REASONABLE FAILURES**
- **REGULAR BACKUPS DEPENDING ON CRITICAL NATURE OF DATA**
- **PROCEDURES TO CONTINUE BUSINESS DURING FAILURE**

### **ADEQUATE DATA RECOVERY**

- **RECOVER FROM FAILURE AS QUICKLY AND COMPLETELY AS POSSIBLE**
- **PROCEDURES TO BRING DATA RESOURCE CURRENT WITH THE BUSINESS**

### **PROTECTED PRIVACY AND CONFIDENTIALITY**

- **PROTECT PERSON / ORGANIZATION'S RIGHT TO PRIVACY / CONFIDENTIALITY**
- **CONSTANTLY TRACK CURRENT LAWS AND REGULATIONS**

### **APPROPRIATE DATA USE**

- **ONGOING REVIEW OF THE USE OF DATA**



# ***DATA RESPONSIBILITY***

## ◆ **ADEQUATE DATA RESPONSIBILITY**

**DEFINE FORMAL RESPONSIBILITIES FOR MANAGING A SHARED DATA RESOURCE.**

## ◆ **GOOD PRACTICES**

### **AUTHORIZED DATA STEWARDSHIP**

**DATA STEWARD - A PERSON WHO WATCHES OVER THE DATA AND IS RESPONSIBLE FOR THE WELFARE OF THE DATA RESOURCE AND ITS SUPPORT TO THE BUSINESS INFORMATION DEMAND**

#### ■ **STRATEGIC LEVEL (EXECUTIVE LEVEL)**

**LEGAL / FINANCIAL RESPONSIBILITY FOR A MAJOR SEGMENT OF THE DATA RESOURCE  
SETS DIRECTIONS, ESTABLISHES PRIORITIES, COMMITS RESOURCES  
COLLECTIVELY MANAGE THE ORGANIZATION'S DATA RESOURCE**

#### ■ **DETAIL LEVEL (KNOWLEDGE WORKER)**

**KNOWLEDGEABLE ABOUT THE DATA BY REASON OF INTIMATE FAMILIARITY WITH DATA  
ARCHITECTURAL RESPONSIBILITY  
DEVELOPS DATA ARCHITECTURE AND DATA RESOURCE DATA  
NO DECISION MAKING AUTHORITY**

#### ■ **TACTICAL LEVEL (LIAISON)**

**LARGE OR MULTI-NATIONAL ORGANIZATIONS FACE DATA GLOBALIZATION ISSUES  
NEED ADDITIONAL LEVEL OF STEWARDSHIP  
REPRESENT UNIQUENESS OF INDIVIDUAL DEPARTMENTS OR COUNTYS**



# ***DATA RESPONSIBILITY***

## **REASONABLE MANAGEMENT PROCEDURES**

- DEVELOPED BY DATA STEWARDS
- EQUIVALENT TO FINANCIAL, HUMAN RESOURCE, PROPERTY MANAGEMENT
- MUST BE REASONABLE - ENCOURAGE PARTICIPATION
- MUST BE READILY AVAILABLE AND UNDERSTANDABLE
  
- PROCEDURES YOU WOULD LIKE TO SEE IF YOU WERE ON THE RECEIVING END.

## **CENTRALIZED CONTROL**

- RESULT FROM DATA STEWARDS AND MANAGEMENT PROCEDURES
- SINGLE COMMON DATA ARCHITECTURE WITH LOCAL IMPLEMENTATION
- NOT A SINGLE 'DATABASE IN THE SKY'
  
- REVERSE OF TRADITIONAL APPROACH
  - MANDATE CENTRALIZED CONTROL
  - ESTABLISH PROCEDURES
  - STEWARDS TO ENFORCE PROCEDURES

***BEGIN THINKING GLOBALLY AND ACTING LOCALLY!***



# ***DATA VISION***

## ◆ ***EXPANDED DATA VISION***

***AN INTELLIGENT FORESIGHT ABOUT THE DATA RESOURCE THAT INCLUDES THE SCOPE OF THE DATA RESOURCE, ITS DEVELOPMENT DIRECTION, AND A PLANNING HORIZON.***

## ◆ ***GOOD PRACTICES***

### ***WIDER SCOPE***

**MUST ULTIMATELY INCLUDE ALL DATA AT AN ORGANIZATION'S DISPOSAL**

#### ■ **NON-CRITICAL DATA**

**HAVE SOME BUSINESS USEFULNESS - MUST ULTIMATELY BE MANAGED**

#### ■ **NON-TABULAR DATA**

***DATA MEGATYPE* - MAJOR GROUPING OF DATA BASED ON OVERALL STRUCTURE AND PHYSICAL MANAGEMENT, SUCH AS SPATIAL, IMAGE, TEXTUAL, ETC.**

***BASE DATA TYPE* - SPECIFIC FORM OF DATA WITHIN A DATA MEGATYPE SUCH AS DATE, TEXT, INTEGER, ETC.**

***DISTINCT DATA TYPE* - DATA UNIT OR STYLE WITHIN A BASE DATA TYPE SUCH AS METERS / YARDS / INCHES OR POUNDS / MARKS / US DOLLARS**

**TABULAR / NON-TABULAR DATA OFTEN CONFUSED WITH  
STRUCTURED / UNSTRUCTURED DATA!**



# ***DATA VISION***

- **HISTORICAL DATA**  
PROPERLY MANAGE TEMPORAL OR 'TIME RELATIONAL' DATA

***OPERATIONAL DATA*** - SUBJECT ORIENTED, INTEGRATED, TIME CURRENT, VOLATILE DATA  
IN SUPPORT OF DAILY OPERATIONS AND OPERATIONAL DECISIONS.

***EVALUATIONAL DATA*** - SUBJECT ORIENTED, INTEGRATED, TIME VARIANT, NON-VOLATILE DATA  
IN SUPPORT OF MANAGEMENT DECISION MAKING.

- **NON-AUTOMATED DATA**  
INCLUDE IN SAME PRIORITIZATION AS AUTOMATED DATA  
OPENS UP WHOLE NEW HIDDEN DATA RESOURCE

## **REASONABLE DEVELOPMENT DIRECTION**

- **PRIMARILY ON BUSINESS DIRECTION**
- **SECONDARILY ON DATABASE MANAGEMENT SYSTEM DIRECTION**

## **REALISTIC PLANNING HORIZON**

- **CHALLENGING YET ACHIEVABLE**
- **STRETCH THE IMAGINATION BUT NOT BE UNREALISTIC**
- **MUST BE UNDERSTANDABLE AND ACHIEVABLE**
- **DIFFERENT HORIZONS FOR DIFFERENT AUDIENCES**

## **COOPERATIVE ESTABLISHMENT**

- **ESTABLISHED AND ACCEPTED COLLECTIVELY BY THE STAKEHOLDERS**



# ***DATA RECOGNITION***

## ◆ **APPROPRIATE DATA RECOGNITION**

**RECOGNIZE THAT DATA ARE A CRITICAL RESOURCE AND IMPROVE THE QUALITY OF THAT CRITICAL RESOURCE TO MEET THE BUSINESS INFORMATION DEMAND**

## ◆ **GOOD PRACTICES**

### **TARGET VESTED INTEREST**

- **HIGHEST LEVEL IN ORGANIZATION WITH VESTED INTEREST IN DATA RESOURCE**
- **WORK FROM THERE THROUGH THE ORGANIZATION**
- **THIS IS HIGHEST PROBABILITY OF SUCCESS**
- **LEVEL VARIES FROM ONE ORGANIZATION TO ANOTHER**

### **DIRECT BUSINESS INVOLVEMENT**

- **DIRECT INVOLVEMENT OF KNOWLEDGEABLE PEOPLE**
- **THE GREATER THE LIMITATION OF BUSINESS -- THE GREATER THE CHANCE OF FAILURE**

***THE GREATER THE INVOLVEMENT OF BUSINESS  
THE GREATER THE CHANCE OF SUCCESS!***

### **TAP THE KNOWLEDGE BASE**

- **FIND PEOPLE WITH KNOWLEDGE OF THE DATA -- BUSINESS OR INFORMATION TECHNOLOGY**
- **USE THEM AS DETAIL DATA STEWARDS OR CONTRIBUTORS**



# ***DATA RECOGNITION***

## **START WITHIN CURRENT BUDGET**

- CURRENT BUDGET INITIATIVES GET RECOGNIZED
- START SMALL AND GROW AS THE RECOGNITION GROWS

## **INCREMENTALLY COST EFFECTIVE APPROACH**

- LONG-TERM BENEFITS HAS LOST ITS MEANING
- SHORT-TERM BENEFITS LOSING ITS MEANING
- *INCREMENTALLY COST EFFECTIVE* -- LAY A SUCCESS ON THE TABLE, GET MORE RESOURCES  
LAY ANOTHER SUCCESS ON THE TABLE, AND SO ON.

## **PROOF POSITIVE PERSPECTIVE**

- GO TO MANAGEMENT WITH A PROOF-POSITIVE PERSPECTIVE
- GAIN RESOURCES TO CONTINUE SUCCESS
- FAR BETTER THAN A 'TRUST ME' REQUEST FOR RESOURCES

***THERE IS A HIGHER RISK -- BUT THE REWARDS FAR OUTWEIGH THE RISKS!***

## **BE OPPORTUNISTIC**

- SEIZE EVERY OPPORTUNITY TO DELIVER THE MESSAGE
- SIZE OF OPPORTUNITY NOT IMPORTANT -- GETTING THE MESSAGE OUT IS IMPORTANT
- AVOID EXCESSIVE PILOT STUDIES AND PROOF OF CONCEPTS
- AVOID PRECISE RIGID PLANS



# ***DATA RECOGNITION***

## **BUILD ON LESSONS LEARNED**

- EVERY PROJECT HAS SUCCESSES AND FAILURES
- STOPPING DATA DISPARITY IS A DISCOVERY PROCESS
- MANAGEMENT MUST ACCEPT CONSTRUCTIVE FAILURES

## **NO BLAME - NO WHITEWASH ATTITUDE**

- THE CURRENT SITUATION EXISTS -- THAT IS A FACT
- BLAME POLARIZES AND ALIENAGES PEOPLE -- DEFEATS THE INITIATIVE
- COVERING IT UP DELAYS ANY INITIATIVE FOR IMPROVEMENT

## **NO UNNECESSARY JUSTIFICATION**

- ELIMINATE THE REQUIREMENT FOR EXTENSIVE JUSTIFICATION TO IMPROVE QUALITY
- THE NEED FOR IMPROVED DATA RESOURCE QUALITY SHOULD BE A TRUISM

***IF ANYTHING -- A JUSTIFICATION SHOULD BE REQUIRED TO CONTINUE  
RUINING THE QUALITY OF A CRITICAL RESOURCE!!***



# ***DATA RESOURCE TRANSITION***

## ◆ ***DATA RESOURCE TRANSITION***

***FORMALLY MOVING FROM A DISPARATE DATA RESOURCE TO A COMPARATE DATA RESOURCE WITHIN THE COMMON DATA ARCHITECTURE***

- RESOLVES THE EXISTING DATA DISPARITY
- NOT A MIGRATION - IMPLIES WANDERING OR PERIODIC RETURN TO DISPARITY
- TRANSITION THROUGH FORMAL TRANSITION STATES

## ◆ ***DATA RESOURCE TRANSITION PROCESS***

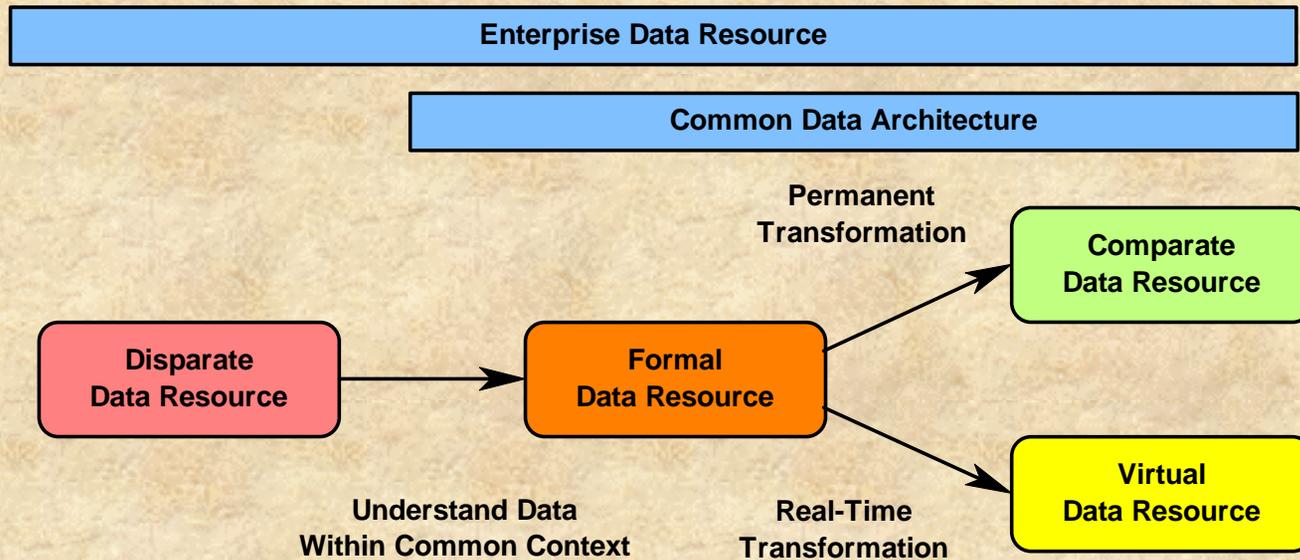
- DISCOVERY PROCESS
- REQUIRES THOUGHT, ANALYSIS, INTUITION, PERCEPTION, AND SOME LUCK
- EVOLUTIONARY PROCESS DUE TO UNCERTAINTY
  
- FORMALIZETHE UNDERSTANDING OF DISPARATE DATA
- INTEGRATE DISPARATE DATA
  
- AVOID *SUCK-AND-SQUIRT* APPROACHES
- AVOID *BRUTE-FORCE-PHYSICAL* APPROACHES

***DATA RESOURCE TRANSITION IS A FORMAL PROCESS!***



# DATA RESOURCE TRANSITION

## ◆ DATA RESOURCE TRANSITION STATES



- **DISPARATE DATA RESOURCE - CURRENT STATE**
  - DATA NOT WELL UNDERSTOOD
- **FORMAL DATA RESOURCE - NECESSARY STATE**
  - NECESSARY TO THOROUGHLY UNDERSTAND DISPARATE DATA
- **VIRTUAL DATA RESOURCE - DESIRED STATE**
  - REAL-TIME PRODUCTION TRANSFORMATION OF DISPARATE DATA
- **COMPARE DATA RESOURCE - IDEAL STATE**
  - PERMANENT TRANSFORMATION OF DISPARATE DATA



# ***FORMAL DATA RESOURCE***

## **◆ FORMAL DATA RESOURCE**

- **DATA INVENTORY**
  - INVENTORY EXISTING DISPARATE DATA
  - RAISES AWARENESS OF THE DATA THAT EXIST
  - SOLVES THE FIRST PROBLEM WITH DISPARATE DATA
  
- **DATA CROSS REFERENCE**
  - CROSS-REFERENCE DISPARATE DATA TO COMMON DATA ARCHITECTURE
  - PROVIDES AN UNDERSTANDING OF DISPARATE DATA IN A COMMON CONTEXT
  - SOLVES THE SECOND PROBLEM WITH DISPARATE DATA
  
- **DATA VARIABILITY**
  - IDENTIFY DATA VARIABILITY
  - DESIGNATE PREFERRED DATA VARIATIONS
  - DEVELOP DATA TRANSLATION SCHEMES
  - PREPARATION TO SOLVE THE THIRD PROBLEM WITH DISPARATE DATA
  
- **DATA REDUNDANCY**
  - IDENTIFY DATA REDUNDANCY
  - DESIGNATE PREFERRED DATA SOURCES
  - PREPARATION TO SOLVE THE FOURTH PROBLEM WITH DISPARATE DATA

***DEVELOPING A FORMAL DATA RESOURCE IS NON-DESTRUCTIVE!***



# ***FORMAL DATA RESOURCE***

## **◆ ENTERPRISE DATA ARCHITECTURE**

- **REMEMBER - THE ENTERPRISE DATA ARCHITECTURE REPRESENTS THE COMPARETE DATA RESOURCE**
- **THE PREFERRED DESIGNATIONS DEFINE THE COMPARETE DATA RESOURCE**
  - **PREFERRED DATA SOURCES**
  - **PREFERRED DATA VARIATIONS**
  - **PREFERRED DATA INTEGRITY RULES**
  - **DATA TRANSLATION SCHEMES**
- **THESE SET THE STAGE FOR TRANSFORMING DISPARTE DATA TO COMPARETE DATA**

***THE PREFERRED DESIGNATIONS DEFINE THE ENTERPRISE DATA ARCHITECTURE!***



# ***DATA TRANSFORMATION***

## ◆ ***PROBLEMS***

- MANY ORGANIZATIONS NOT FORMALLY INTEGRATING THEIR DATA
- USUALLY INCOMPLETE / INCONSISTENT DATA INTEGRATION
- IDENTIFYING A SINGLE SYSTEM OF REFERENCE USUALLY NOT VALID
- USING *SUCK-AND-SQUIRT* METHODS TO OBTAIN DATA USUALLY DON'T WORK
- USING *BRUTE-FORCE-PHYSICAL* APPROACHES
- MORE THAN SIMPLE EXTRACT - TRANSFORM - LOAD
- MORE THAN MINOR DATA CLEANISING
- RESULT IS LOW-QUALITY DATA

## ◆ ***APPROACH***

- FORMAL DATA TRANSFORMATION WITHIN COMMON DATA ARCHITECTURE
- BASED ON FORMAL DATA RESOURCE - THOROUGH UNDERSTANDING OF DATA
- REQUIRES ROBUST DATA TRANSFORMATION
- NOT A TRIVIAL TASK - BUT NOT IMPOSSIBLE EITHER
- IGNORED IN MANY METHODS

***DATA INTEGRATION MUST BE BASED ON THE FORMAL DATA RESOURCE!***



# ***DATA TRANSFORMATION***

## ◆ ***VIRTUAL / COMPARETE DATA RESOURCE***

### ● **EXTRACT - THREE STEPS**

- **IDENTIFYING AND EXTRACTING THE BEST DISPARATE DATA TO A DATA DEPOT**
- **ISOLATED FROM SOURCE AND TARGET DATA**
- **PREPARATION FOR DATA TRANSFORMATION**

### ● **TRANSFORM - FIVE STEPS**

- **TRANSFORMING THE DISPARATE DATA TO COMPARETE DATA**
- **USING FORMAL DATA TRANSFORMATION RULES**
- **ACTUALLY CHANGING THE DATA**

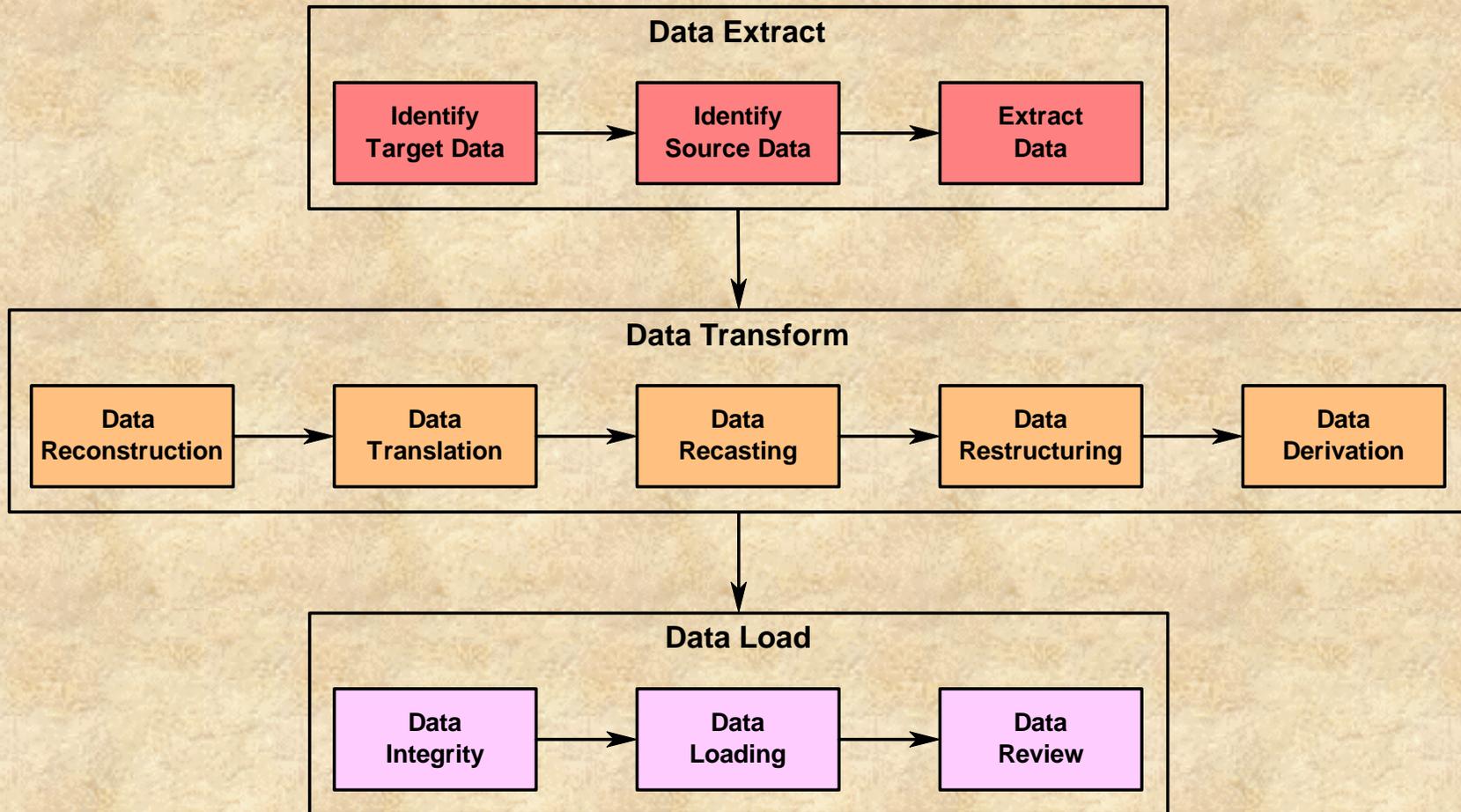
### ● **LOAD - THREE STEPS**

- **EDITING THE COMPARETE DATA**
- **LOADING THE COMPARETE DATA FROM A DATA DEPOT**
- **BUILDING THE COMPARETE DATA RESOURCE**



# DATA TRANSFORMATION

## ◆ DATA TRANSFORMATION PROCESS





# ***CULTURAL CONSIDERATIONS***

## **◆ A SHARABLE DATA RESOURCE**

- **GETTING PEOPLE TO INTERACT AND SHARE**
  - **AWARE OF EXISTING DATA**
  - **THOROUGHLY UNDERSTAND DATA**
  - **USE DATA TO SUPPORT THEIR BUSINESS ACTIVITIES**
  
- **REAL DATA SHARING INCENTIVES**
  - **MAKE EASY TO MEET BUSINESS NEEDS**
  - **MAKE EASY TO CROSS ORGANIZATIONAL BOUNDARIES**

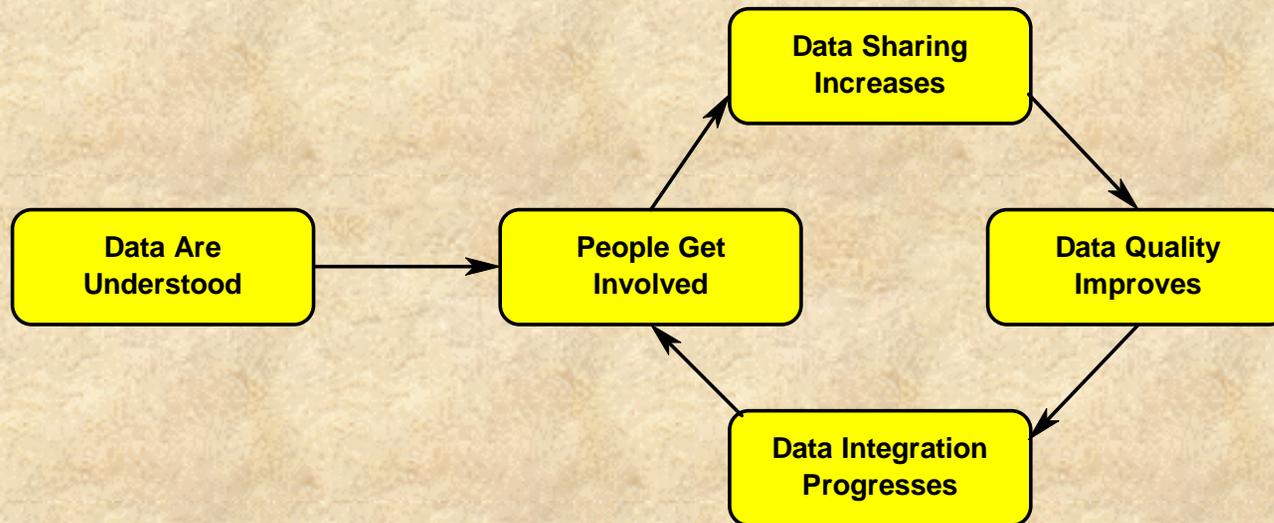
***WHEN THE PAIN OF NOT MEETING BUSINESS NEEDS  
EXCEEDS THE PAIN OF CROSSING ORGANIZATIONAL BOUNDARIES  
PEOPLE WILL INTERACT AND SHARE!***



## ***CULTURAL CONSIDERATIONS***

### ◆ ***DATA SHARING - DATA QUALITY - DATA INTEGRATION CYCLE***

- **WHEN PEOPLE UNDERSTAND DATA - THEY SHARE THOSE DATA**
- **WHEN PEOPLE SHARE DATA - THE DATA QUALITY IMPROVES**
- **WHEN DATA QUALITY IMPROVES - THE DATA ARE INTEGRATED**
- **AND THE CYCLE KEEPS GOING**



***THIS IS NOT ROCKET SCIENCE!***



# ***CULTURAL CONSIDERATIONS***

## **◆ A PEOPLE ORIENTATION**

- **UNCERTAINTY**
  - HELP RESOLVE UNCERTAINTY
  - REDUCE RESISTANCE TO CHANGE
  - MANAGE EXPECTATIONS
  
- **WILLINGNESS TO CHANGE**
  - MOST PEOPLE WILLING TO CHANGE
  - THEY JUST DON'T KNOW HOW
  - CONCERNED ABOUT IMPACT TO BUSINESS
  
- **RESISTANCE TO CHANGE**
  - PEOPLE UNSURE OF NEW PARADIGM
  - COMFORTABLE WITH EXISTING PARADIGM
  - CONCERNED ABOUT PAST CURE-ALL FAILURES
  - NO TIME TO STOP THE BUSINESS TO CHANGE
  
- **SURFACE DIFFERENCES**
  - DOCUMENT DIFFERENT DEFINITIONS
  - DON'T ATTEMPT TO RESOLVE DIFFERENCES

***PEOPLE DON'T MIND CHANGING!  
PEOPLE DO MIND BEING CHANGED!***



# ***CULTURAL CONSIDERATIONS***

## **◆ A BUSINESS ORIENTATION**

- **FOCUS ON THE BUSINESS**
  - **BASED ON BUSINESS PLAN AND STRATEGIES**
  - **BASED ON ENTERPRISE'S PERCEPTION OF BUSINESS WORLD**
  
- **CRITICAL AREA APPROACH**
  - **IDENTIFY CRITICAL BUSINESS AREAS**
  - **DETERMINE DATA NEEDED FOR THOSE AREAS**
  - **IDENTIFY DATA THAT ALREADY EXIST**
  - **IDENTIFY DATA THAT NEED TO BE ACQUIRED**
  - **DEVELOP COMPARETE DATA FOR THOSE CRITICAL AREAS**
  
- **CLIENT FOCUSED APPROACH**
  - **ORIENT THE PROCESS TO THE CLIENT'S WAY OF THINKING**
  - **DRAW THE CLIENT INTO THE PROCESS**
  - **TAP THE CLIENT'S KNOWLEDGE BASE**
  
- **DIRECT CLIENT INVOLVEMENT**
  - **INVOLVEMENT LEADS TO COMMITMENT**
  - **COMMITMENT LEADS TO ACCEPTANCE**
  - **ACCEPTANCE LEADS TO SUCCESS**



# ***CULTURAL CONSIDERATIONS***

## **◆ SUCCESS MOTIVATION**

- **SUCCESS IS CONTAGIOUS**
  - ORIENT TOWARD CONTINUED SUCCESSES
  - THRIVE ON LESSONS LEARNED
- **INCREMENTALLY COST EFFECTIVE**
  - LONG TERM BENEFITS HAS LOST ITS PUNCH
  - SHORT TERM BENEFITS BEGINNING TO LOSE ITS PUNCH
  - INCREMENTALLY COST EFFECTIVE IS A GOOD TERM
- **AVOID MANDATES**
  - MANDATES DO NOT WORK
  - IMPLY COMPLIANCE, ENFORCEMENT, PUNISHMENT
  - CREATES A WIN-LOSE SITUATION
- **ESTABLISH GUIDELINES**
  - GAIN CONSENSUS AND ACCEPTANCE
  - EVOLVE WITH TECHNOLOGY
- **DESIRED APPROACH**
  - EASIEST ROUTE TO FOLLOW
  - ASSURE SUCCESS

***CREATE WIN - WIN SITUATION!***



## ***CONCLUSION***

### ◆ ***ULTIMATE DATA RESOURCE QUALITY IS STABILITY ACROSS***

- CHANGING BUSINESS STRATEGIES & INFORMATION NEEDS
- CHANGING TECHNOLOGY

### ◆ ***STABILITY ACHIEVED BY***

- ESTABLISHING A COMMON DATA ARCHITECTURE
- HALTING THE CONTINUED DATA DISPARITY
- RESOLVING EXISTING DATA DISPARITY

### ◆ ***NO SILVER BULLETS***

- ATTEMPT TO ACHIEVE SOME GAIN WITHOUT ANY PAIN
- USUALLY ENDURE CONSIDERABLE PAIN WITH MINIMAL GAIN

### ◆ ***DON'T DELAY***

- NEED NO JUSTIFICATION TO IMPROVE DATA RESOURCE QUALITY
- CONCEPTS, PRINCIPLES, & TECHNIQUES ARE AVAILABLE
- THE SITUATION IS ONLY GETTING WORSE



## ***PRESENTER'S BIBLIOGRAPHY***

***DEVELOPING DATA STRUCTURED INFORMATION SYSTEMS***

**KEN ORR & ASSOCIATES, 1984**

***DEVELOPING DATA STRUCTURED DATABASES***

**PRENTICE HALL, 1987**

***PRACTICAL DATA DESIGN***

**PRENTICE HALL, 1990**

***DATA SHARING USING A COMMON DATA ARCHITECTURE***

**JOHN WILEY & SONS, 1994**

***THE DATA WAREHOUSE CHALLENGE: TAMING DATA CHAOS***

**JOHN WILEY & SONS, 1996**

***DATA RESOURCE QUALITY: TURNING BAD HABITS INTO GOOD PRACTICES***

**ADDISON WESLEY, 2000**

***DATA RESOURCE INTEGRATION: UNDERSTANDING & RESOLVING DATA DISPARITY***

**IN PROGRESS, NO PUBLISHER OR DUE DATE**

***DATA RULES: KEY TO DATA RESOURCE QUALITY***

**IN PROGRESS, NO PUBLISHER OR DUE DATE**

**CONSULT PRESENTER'S WEB PAGE FOR ADDITIONAL DETAILS**