



***CALIFORNIA LONGITUDINAL PUPIL  
ACHIEVEMENT DATA SYSTEM:***

***BUILDING BLOCKS FOR  
SUCCESSFUL IMPLEMENTATION***

***PART 1  
COMMON DATA ARCHITECTURE***

***NCES MIS CONFERENCE  
FEBRUARY 28, 2008***

***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



# ***THE PROBLEM***

## ◆ ***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***

## ◆ ***BUSINESS INFORMATION DEMAND IS NOT BEING MET***

- **CURRENT BUSINESS INFORMATION DEMAND**
- **FUTURE BUSINESS INFORMATION DEMAND**

## ◆ ***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 PROBLEM***

- **UNKNOWN DATA EXISTENCE**
- **UNKNOWN DATA MEANING**
- **HIGH DATA REDUNDANCY**
- **HIGH DATA VARIABILITY**

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



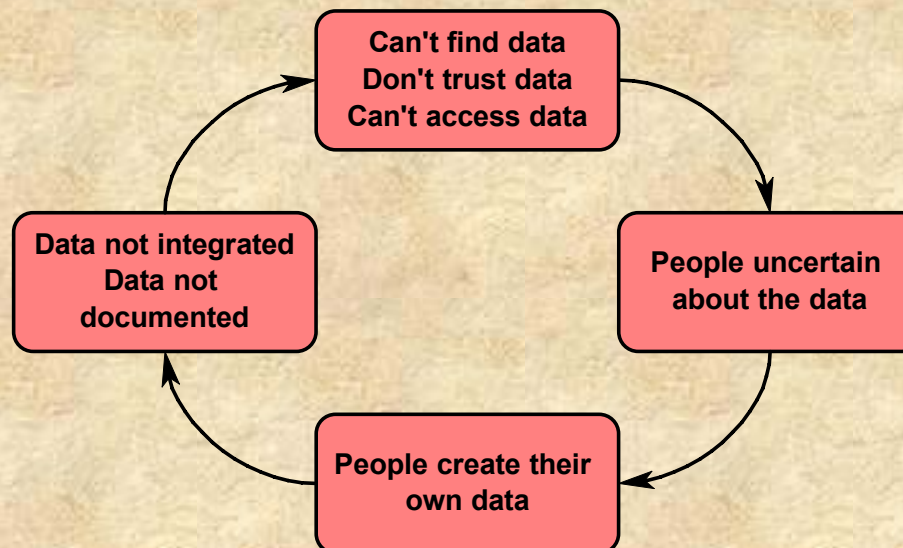
## ***THE PROBLEM***

### ◆ ***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!***





## ***THE SOLUTION***

### ◆ ***DATA RESOURCE NATURALLY DRIFTS TOWARD DISPARITY***

- DRIFTS TOWARD HIGHER ENTROPY
- ACCORDING TO SECOND LAW OF THERMODYNAMICS
- WILL NOT REACH AN EQUILIBRIUM – AN OPEN SYSTEM
- WILL NOT IMPROVE ON ITS OWN

### ◆ ***MUST CONSCIOUSLY PREVENT AND RESOLVE DATA DISPARITY***

- PHASE 1  
STOPPING THE CONTINUED CREATION OF DISPARATE DATA
  
- PHASE 2  
RESOLVING EXISTING DATA DISPARITY  
CREATING A HIGH-QUALITY SHARABLE DATA RESOURCE

### ◆ ***ORGANIZATIONS MUST BEGIN AN INITIATIVE***

***THOROUGHLY UNDERSTAND , FORMALLY MANAGE , AND FULLY UTILIZE ALL DATA THAT ARE AVAILABLE TO THE ORGANIZATION WITHIN ONE COMMON ORGANIZATION-WIDE DATA ARCHITECTURE***

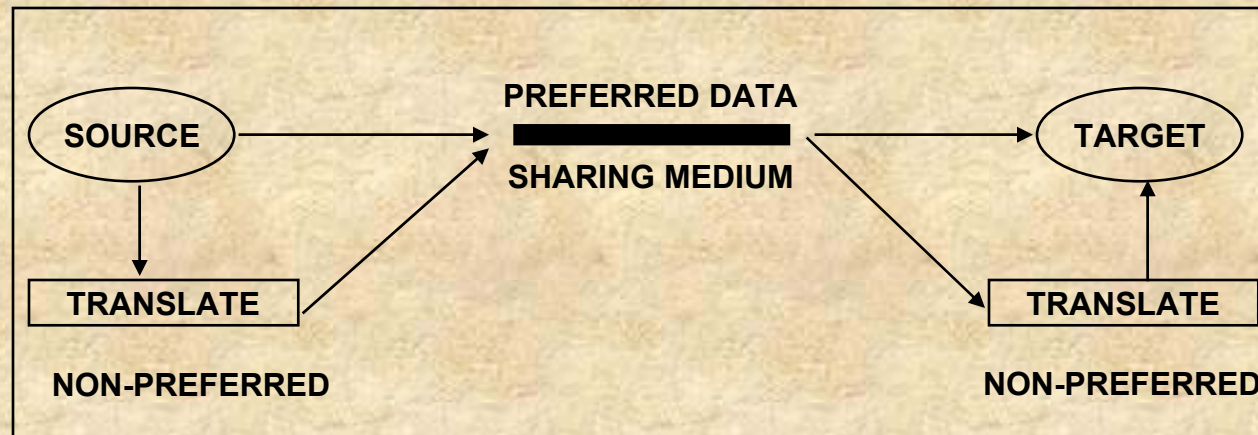
***ORGANIZATIONS MUST TAKE THE INITIATIVE NOW!***



# ***THE SOLUTION***

## ◆ **DATA SHARING VISION**

**UNDERSTAND DATA IN A COMMON CONTEXT SO THEY CAN BE READILY SHARED TO MEET THE CURRENT AND FUTURE BUSINESS INFORMATION DEMAND**



- **DATA SHARED THROUGH PREFERRED VARIATIONS**
- **SOURCE ORGANIZATION TRANSLATES TO PREFERRED IF NECESSARY**
- **TARGET TRANSLATIONS TRANSLATES FROM PREFERRED IF NECESSARY**
- **SOURCE AND TARGET CAN CHANGE INDEPENDENTLY**

***DATA SHARING DONE WITH PREFERRED DATA!***



# ***THE SOLUTION***

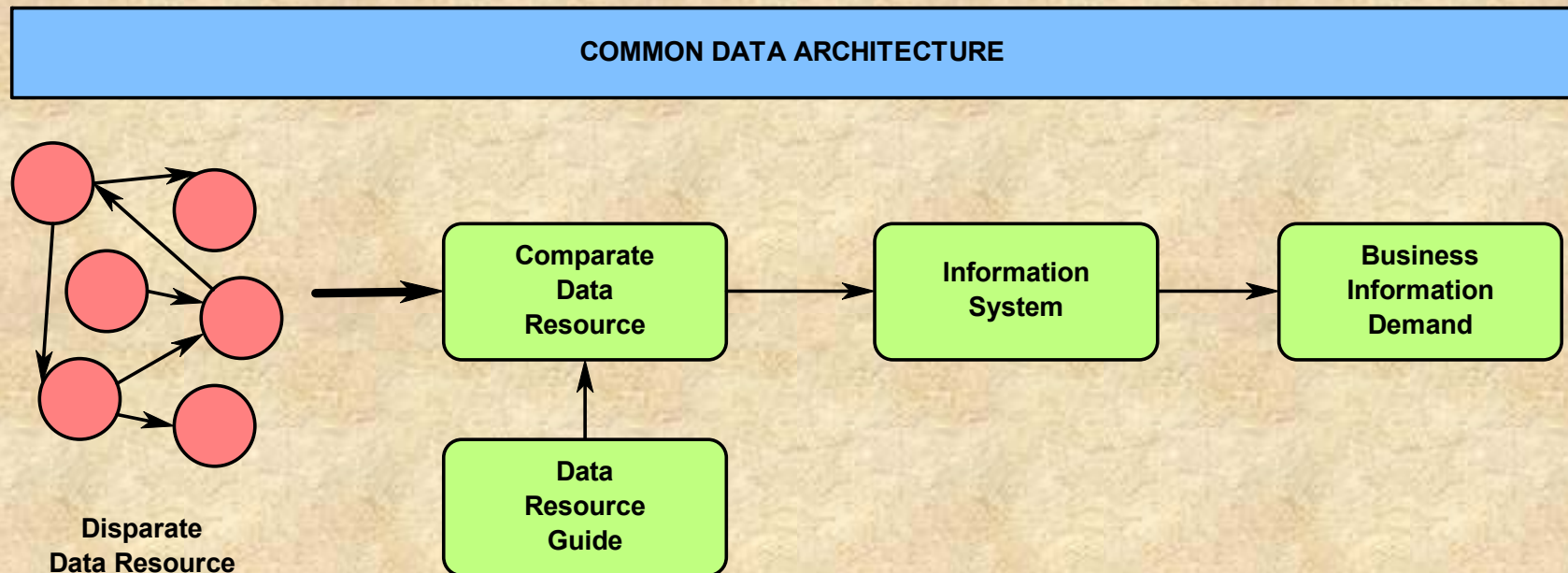
## ◆ **COMPARETE DATA VISION**

- **COMPARETE DATA**

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

- **COMPARETE DATA RESOURCE**

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

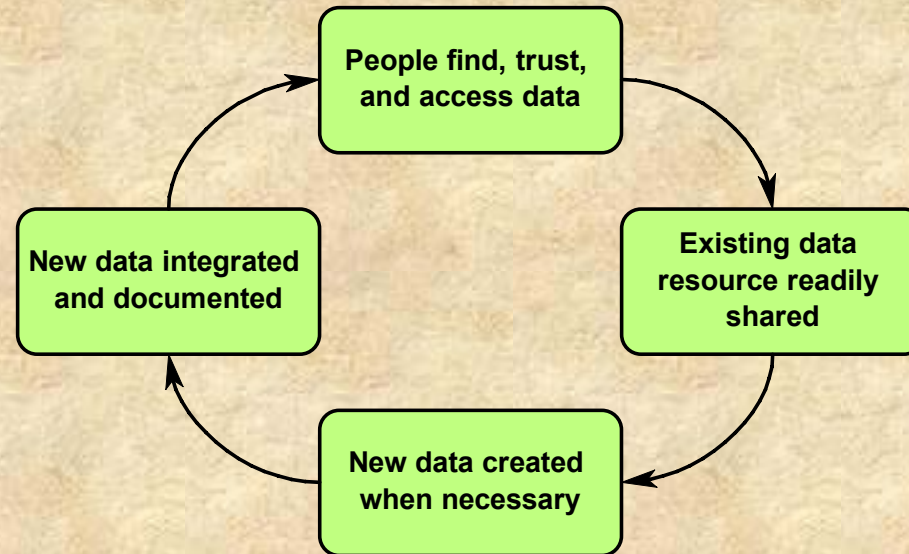




## ***THE SOLUTION***

### ◆ **COMPARETE DATA CYCLE**

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



***HALT THE DISPARATE DATA CYCLE  
AND START A COMPARTE DATA CYCLE!***





## ***THE SOLUTION***

### ◆ ***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***

***BUILDING A SHARABLE DATA RESOURCE REQUIRES A FORMAL DEFINITION OF DATA ARCHITECTURE!***





# ***THE SOLUTION***

## ◆ ***COMMON DATA ARCHITECTURE***

- **A DATA ARCHITECTURE AS DEFINED**
- **PLUS A COMMON CONTEXT FOR**
  - **INVENTORYING ALL DATA**
  - **UNDERSTANDING THE CONTENT, MEANING, AND INTEGRITY OF ALL DATA**
  - **INTEGRATING DISPARATE DATA**
  - **IMPROVING DATA QUALITY**
  - **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**

***THE COMMON DATA ARCHITECTURE TRANSCENDS ALL DATA!***



# ***STOPPING DATA DISPARITY***

## **◆ BASIC ARCHITECTURE CONCEPTS**

### **● 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 WITH A FORMALS ET OF ABBREVIATIONS AND AN ABBREVIATION ALGORITHM***

### **● 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***

### **● PROPER DATA STRUCTURES**

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

### **● PRECISE DATA INTEGRITY RULES**

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

### **● ROBUST DATA DOCUMENTATION**

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



# ***STOPPING DATA DISPARITY***

## **◆ BASIC GOVERNANCE CONCEPTS**

### **● REASONABLE DATA ORIENTATION**

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

### **● ACCEPTABLE DATA AVAILABILITY**

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

### **● ADEQUATE DATA RESPONSIBILITY**

***DEFINE FORMAL DATA STEWARD RESPONSIBILITIES FOR MANAGING A SHARED DATA RESOURCE, INCLUDING STRATEGIC, TACTICAL, AND DETAIL DATA STEWARDS***

### **● 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***

### **● APPROPRIATE DATA RECOGNITION**

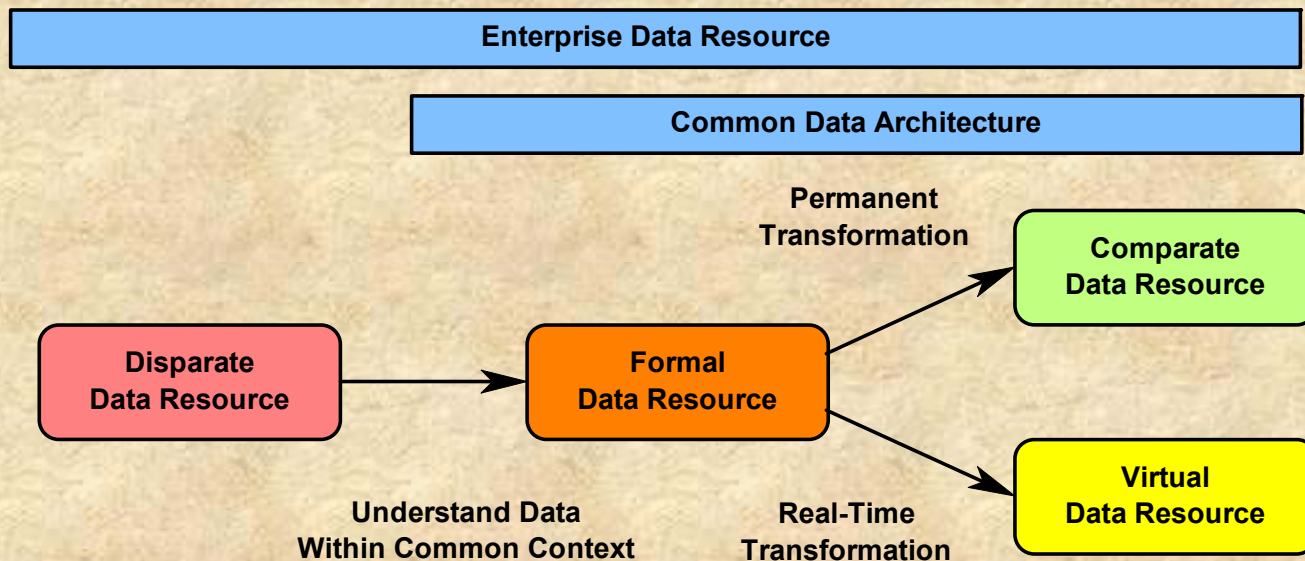
***RECOGNIZE THAT DATA ARE A CRITICAL RESOURCE AND IMPROVE THE QUALITY OF THAT CRITICAL RESOURCE TO MEET THE BUSINESS INFORMATION DEMAND***



# ***UNDERSTANDING DISPARATE DATA***

## **◆ DATA RESOURCE TRANSITION VISION**

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



- **DISPARATE DATA RESOURCE - CURRENT STATE**
- **FORMAL DATA RESOURCE - NECESSARY STATE**
- **VIRTUAL DATA RESOURCE - DESIRED STATE**
- **COMPARE DATA RESOURCE - IDEAL STATE**





# ***UNDERSTANDING DISPARATE DATA***

## **◆ DEVELOP A FORMAL DATA RESOURCE**

**FORMALIZE THE UNDERSTANDING OF DISPARATE DATA WITHIN A COMMON DATA ARCHITECTURE**

- DEVELOPING A FORMAL DATA RESOURCE IS NON-DESTRUCTIVE
- THE CURRENT SYSTEMS CONTINUE TO OPERATE UNAFFECTED
- FIVE SPECIFIC STEPS

## **◆ STEP 1 - DATA INVENTORY**

- IDENTIFY AND DOCUMENT ALL DATA PRODUCTS WITHIN THE CURRENT SCOPE SO THEY CAN BE CROSS REFERENCED TO THE COMMON DATA ARCHITECTURE
- DATA PRODUCTS ARE A PRODUCT OF ANY DEVELOPMENT EFFORT – DATABASE, DICTIONARY, DATA MODEL, APPLICATION, ETC.
- RAISES AWARENESS OF THE EXISTING DISPARATE DATA

***SOLVES THE FIRST PROBLEM WITH DISPARATE DATA!***



# ***UNDERSTANDING DISPARATE DATA***

## **◆ STEP 2 - DEVELOP AN INITIAL COMMON DATA ARCHITECTURE**

- **DEVELOP A HIGH-LEVEL COMMON DATA ARCHITECTURE BASED ON BUSINESS OBJECTS AND BUSINESS EVENTS IN THE REAL WORLD**
  - MAJOR DATA SUBJECTS
  - MAJOR DATA CHARACTERISTICS
  - PRIMARY DATA RELATIONS
  - PRELIMINARY DATA DEFINITIONS
  
- **DO NOT TRY TO DEVELOP THE ENTIRE DESIRED DATA ARCHITECTURE**
  - IT WILL BE A WASTED EFFORT
  - IT WILL LIKELY BE WRONG
  - EXISTING DISPARATE DATA WILL NOT BE UNDERSTOOD
  - CROSS-WALKING WILL BE VERY DIFFICULT
  
- **THE COMMON DATA ARCHITECTURE WILL BE ENHANCED DURING THE CROSS REFERENCING PROCESS**

***STARTS THE COMMON CONTEXT FOR UNDERSTANDING!***



# ***UNDERSTANDING DISPARATE DATA***

## **◆ STEP 3 - CROSS REFERENCE DISPARATE DATA**

- LOGICAL MAPPING OF THE DISPARATE DATA TO COMMON DATA ARCHITECTURE
- UNDERSTAND DISPARATE DATA WITHIN A COMMON CONTEXT
- NOT A CROSS-WALK FOR MOVING DATA; IT'S A CROSS-REFERENCE FOR UNDERSTANDING DATA
- ENHANCE INITIAL COMMON DATA ARCHITECTURE AS NECESSARY TO HANDLE THE CROSS-REFERENCES
- UNDERSTAND EXISTING NAMES AND DEFINITIONS
- UNDERSTAND EXISTING DATA INTEGRITY RULES
- AREA OF DISCOVERY
  - 70 - 80% RELATIVELY EASY
  - 10-15% TAKE SOME THOUGHT
  - 5 - 10% DEEPER INVESTIGATION

***SOLVES THE SECOND PROBLEM WITH DISPARATE DATA!***



# ***UNDERSTANDING DISPARATE DATA***

## **◆ STEP 4 – IDENTIFY DATA REDUNDANCY**

- **IDENTIFY THE REDUNDANCY THAT EXISTS IN DISPARATE DATA**
  - DATA CHARACTERISTICS
  - DATA REFERENCE SETS
  - AVERAGES A FACTOR OF 10 INSTANCES FOR EACH BUSINESS FACT
  
- **DESIGNATE PREFERRED DATA SOURCES – MOST CURRENT AND MOST ACCURATE**
  - ONE PREFERRED SOURCE FOR EACH BUSINESS FACT
  - ONE PREFERRED SOURCE FOR EACH DATA REFERENCE SET
  - NOT A SINGLE SYSTEM OF REFERENCE / SYSTEM OF RECORD
  - CONDITIONAL SOURCING IS THE NORM

***PREPARATION TO SOLVE THE THIRD PROBLEM WITH DISPARATE DATA!***





# ***UNDERSTANDING DISPARATE DATA***

## **◆ STEP 5 - IDENTIFY DATA VARIABILITY**

- IDENTIFY THE VARIABILITY THAT EXISTS WITHIN THE DISPARATE DATA
  
- DESIGNATE PREFERRED VARIATIONS
  - DATA NAMES
  - DATA DEFINITIONS
  - DATA INTEGRITY RULES
  
- THIS IS THE DESIRED DATA ARCHITECTURE – A BY PRODUCT OF THE UNDERSTANDING
  
- DEVELOP DATA TRANSLATION SCHEMES
  - BOTH WAYS BETWEEN PREFERRED AND NON-PREFERRED VARIATIONS
  - THESE WILL BE USED DURING DATA TRANSITION

***PREPARATION TO SOLVE THE FOURTH PROBLEM WITH DISPARATE DATA!***



# ***RESOLVING DISPARATE DATA***

## **◆ DATA TRANSITION**

***THE PROCESS OF TRANSITIONING FROM DISPARATE DATA TO COMPARETE DATA WITHIN THE COMMON DATA ARCHITECTURE***

- NOT A MIGRATION – IT'S A PERMANENT TRANSITION
- FORMAL DATA TRANSITION WITHIN COMMON DATA ARCHITECTURE
- BASED ON FORMAL DATA RESOURCE - THOROUGH UNDERSTANDING OF THE EXISTING DISPARATE DATA

## **◆ BUILT FROM THE DESIRED DATA ARCHITECTURE**

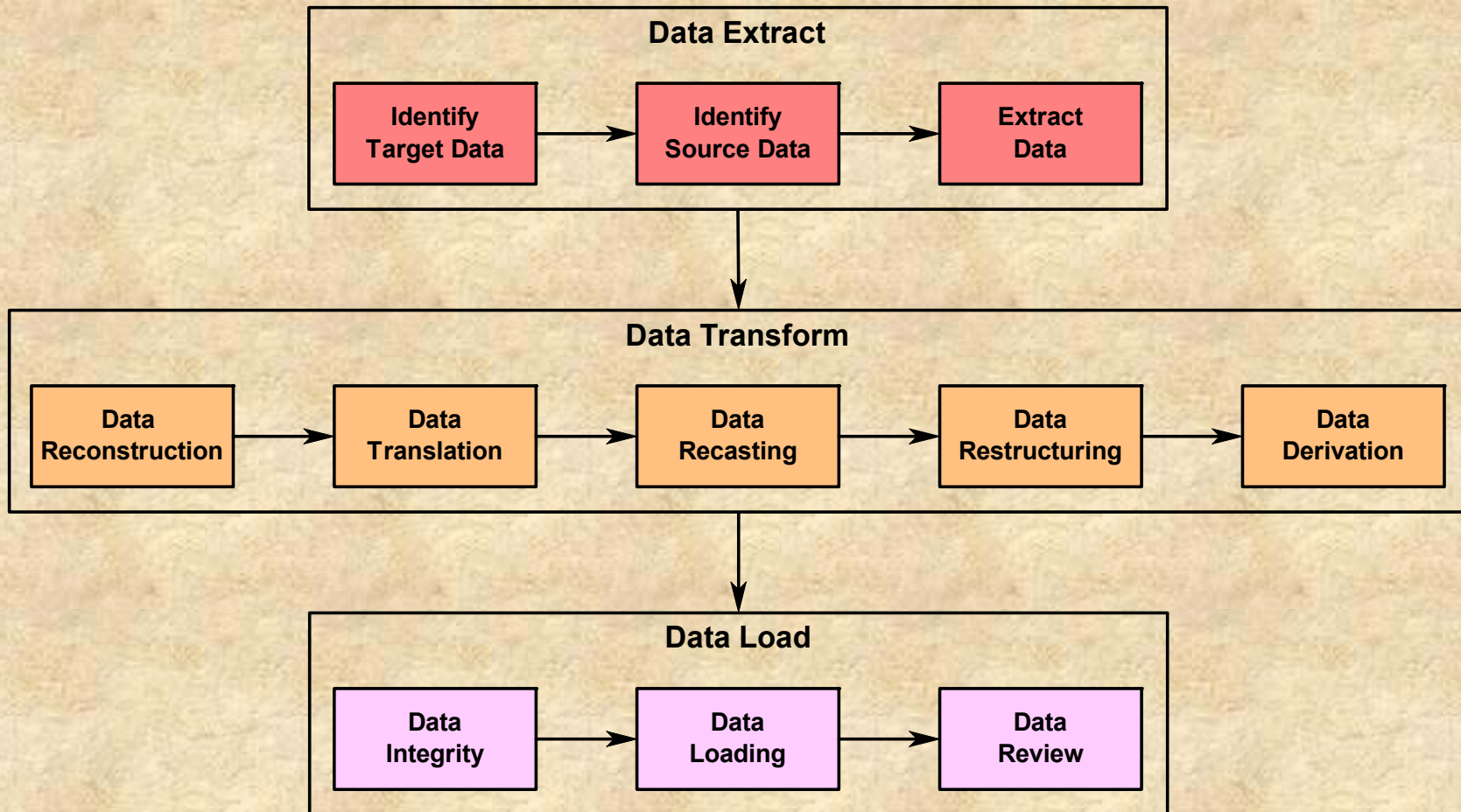
- PREFERRED DATA SOURCES
- PREFERRED DATA NAMES
- PREFERRED DATA DEFINITIONS
- PREFERRED DATA INTEGRITY RULES

***DATA TRANSITION ACTUALLY CHANGES THE DATA!***



# RESOLVING DISPARATE DATA

## ◆ DATA TRANSITION VISION

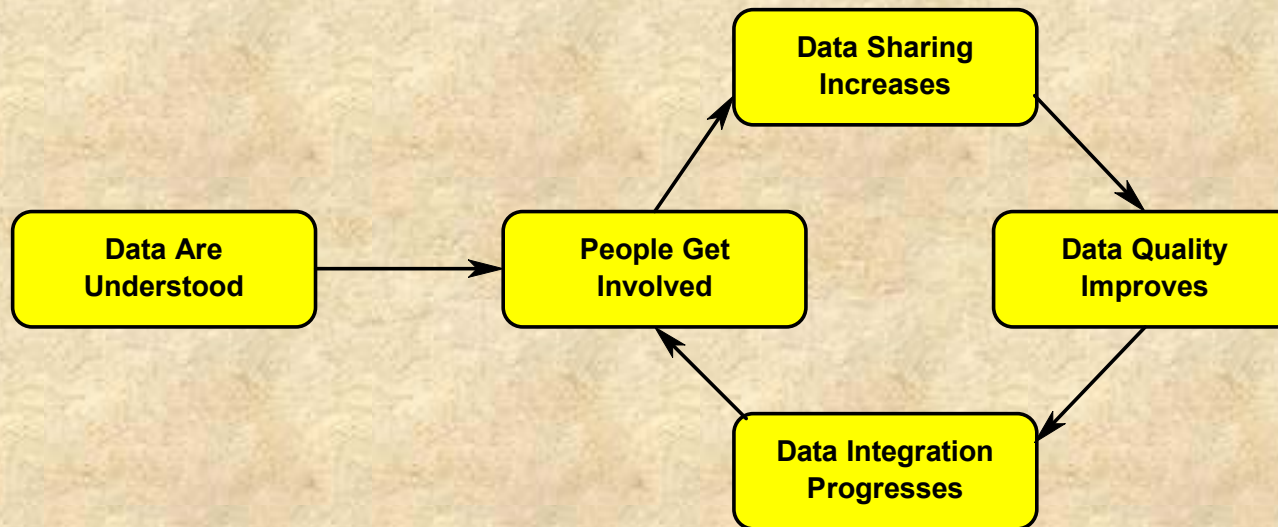




## ***RESULT***

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

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



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





# ***RESULT***

## ◆ ***BENEFITS***

- HIGH QUALITY DATA RESOURCE
- UNDERSTANDABLE DATA RESOURCE
- SHARABLE DATA RESOURCE
- SUPPORT FOR THE CURRENT AND FUTURE BUSINESS INFORMATION DEMAND

## ◆ ***IMPLEMENTATION***

- **LOCALLY DEVELOPED SYSTEMS**
  - DESIGNED ACCORDING TO PREFERRED DATA ARCHITECTURE
- **PURCHASED APPLICATIONS**
  - OPTION 1 - CHANGE THE APPLICATION NAMES, DEFINITIONS, ETC.
  - OPTION 2 - DOCUMENT AS A DATA PRODUCT  
CROSS REFERENCE TO COMMON DATA ARCHITECTURE  
DECIDE HOW TO USE EACH DATA ITEM IN THE APPLICATION

## ◆ ***THREE LEVELS OF A COMMON DATA ARCHITECTURE***

- LOCAL LEVEL
- STATE LEVEL – CURRENT EFFORT
- FEDERAL LEVEL



## ***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, TENTATIVE PUBLISHER: ADDISON-WESLEY**

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

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

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