



*FOUNDATION FOR  
DATA RESOURCE QUALITY*

*CENTRAL OHIO DAMA  
APRIL 14, 2005*

*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



# FOUNDATION

## ◆ OBJECTIVES

- MANY DIFFERENT CONCEPTS AND APPROACHES TO DATA RESOURCE QUALITY
- PUT DIFFERENT CONCEPTS AND APPROACHES IN PERSPECTIVE
- PROVIDE BASIC FOUNDATION FOR DATA RESOURCE QUALITY
- BREAK SOME MYTHS ABOUT DATA RESOURCE QUALITY
- PROVIDE SOME IDEAS ABOUT THE COST OF DATA RESOURCE QUALITY
- STIMULATE THOUGHT ABOUT DATA RESOURCE QUALITY



## ◆ **SECOND LAW OF THERMODYNAMICS**

### ● **IN A CLOSED SYSTEM**

- **ORDERED ENERGY CAN BE CONSUMED ENTIRELY INTO DISORDERED ENERGY**
- **BUT NEVER VISA VERSA**
- **PRINCIPLE OF *INCREASING ENTROPY***
- **SYSTEM REACHES STATE OF MAXIMUM ENTROPY - *AN EQUILIBRIUM***
- **LAW IS INEVITABLE AND IRREVERSIBLE FOR CLOSED SYSTEMS**

### ● **ENTROPY IS THE STATE OR DEGREE OF DISORDERLINESS**

- **ENTROPY IS LOSS OF ORDER**
- **ENTROPY IS INCREASING DISORDERLINESS**
- **ENTROPY INCREASES OVER TIME**
- **THINGS BECOME MORE DISORDERLY OVER TIME**
- **MANY EVERYDAY EXAMPLES WHERE ORDER GIVES WAY TO DISORDERLINESS**

### ● **IN AN OPEN SYSTEM**

- **ENTROPY CONTINUES TO INCREASE**
- **AN EQUILIBRIUM IS NOT REACHED**
- **ENTROPY CAN BE REVERSED WITH INPUT OF ENERGY**
- **MUST DO WORK TO REDUCE ENTROPY AND INCREASE ORDERLINESS**



# FOUNDATION

## ◆ **DATA RESOURCE**

- **THE DATA RESOURCE IS AN *OPEN SYSTEM***
- **FOLLOWS THE PRINCIPLE OF INCREASING ENTROPY**
- **THERE IS NO EQUILIBRIUM – ENTROPY CONTINUES TO INCREASE**
- **BUT THE ENTROPY CAN BE REVERSED WITH INPUT OF ENERGY**

## ◆ **REVERSING ENTROPY**

- **GOING AGAINST THE TIDE**
- **AGAINST THE PRINCIPLE OF INCREASING ENTROPY**
- **CREATING ORDER OUT OF DISORDER**
- **ENERGY IS REQUIRED TO REDUCE ENTROPY AND INCREASE ORDER**
- **THE HIGHER THE ENTROPY, THE MORE ENERGY REQUIRED TO ACHIEVE ORDER**



# FOUNDATION

## ◆ **ENTROPY AND DISPARITY**

- ENTROPY CAN BE EQUATED TO DISPARITY
- ENTROPY IN THE DATA RESOURCE IS EQUATED TO DISPARITY

***ENTROPY = DISORDERLINESS = DISPARITY***

## ◆ **ENERGY AND COST**

- ENERGY CAN BE EQUATED TO EFFORT
- EFFORT CAN BE EQUATED TO COST
- REDUCTION OF ENTROPY / DISPARITY CAN BE BASED ON COST

***ENERGY = EFFORT = COST***



# FOUNDATION

## ◆ **CLAUDE SHANNON**

- **INFORMATION IS THE OPPOSITE OF ENTROPY**
- **ENTROPY IS DISORDERLINESS - *NOISE***
- **INFORMATION MUST BE RELEVANT AND TIMELY**

## ◆ **INFORMATION THEORY**

- **TODAY'S INFORMATION THEORY IS YESTERDAY'S COMMUNICATION THEORY**
- **INTERPERSONAL COMMUNICATION**
- **PROVIDING INFORMATION TO PEOPLE – REGARDLESS OF THE MEDIUM**

## ◆ **THE MEANING OF INFORMATION HAS BEEN CONFUSED**

- ***INFORMATION* USED IN AN UNQUALIFIED MANNER – RELATIVELY USELESS**
  
- ***SYNTACTIC INFORMATION* IS RAW DATA**
- **ARRANGED ACCORDING TO CERTAIN RULES**
- **SYNTACTIC INFORMATION ALONE IS MEANINGLESS – RAW DATA**
  
- ***SEMANTIC INFORMATION* HAS CONTEXT AND MEANING**
- **IT IS RELEVANT AND TIMELY**
- **ALSO ARRANGED ACCORDING TO CERTAIN RULES**



# FOUNDATION

## ◆ *WHAT IS QUALITY*

- THE DEGREE OF ORDERLINESS OF SOMETHING
- HIGHER QUALITY IS A HIGHER LEVEL OF ORDERLINESS – LOWER ENTROPY
- LOWER QUALITY IS A LOWER LEVEL OF ORDERLINESS – HIGHER ENTROPY

## ◆ *ACHIEVING QUALITY*

- HIGHER QUALITY REQUIRES LOWER ENTROPY
- LOWER ENTROPY REQUIRES INPUT OF ENERGY
- SINCE ENERGY IS EQUATED TO COST
- HIGHER QUALITY COMES AT A COST

***QUALITY COMES AT A COST***



## ◆ **DATA RESOURCE QUALITY**

- **TWO ASPECTS OF QUALITY**
  - **SIZE OF THE DATA RESOURCE – HOW MUCH DATA**
  - **LIFE OF THE DATA RESOURCE – HOW LONG IT HAS EXISTED**

***CURRENT QUALITY DEPENDS ON SIZE AND DURATION***

- **QUALITY IMPROVEMENT**
  - **LARGER DATA RESOURCE MEANS MORE ENERGY TO IMPROVE – HIGHER COST**
  - **LONGER LIFE MEANS MORE ENERGY TO IMPROVE – HIGHER COST**

***IMPROVEMENT COST DEPENDS ON CURRENT QUALITY***

## ◆ **INFORMATION QUALITY**

- **TWO ASPECTS OF QUALITY**
  - **QUALITY DEPENDS ON DATA RESOURCE QUALITY**
  - **QUALITY DEPENDS ON INFORMATION PROCESS**

***INFORMATION QUALITY NO BETTER THAN DATA RESOURCE QUALITY***





# FOUNDATION

## ◆ **UNCERTAINTY VS. UNDERSTANDING**

- HIGHER ENTROPY EQUATES TO HIGHER UNCERTAINTY AND LOWER UNDERSTANDING
- AS ENTROPY GOES UP, UNCERTAINTY GOES UP, UNDERSTANDING GOES DOWN
  
- AS THE DATA RESOURCE GETS MORE DISPARATE
- PEOPLE LOSE UNDERSTANDING ABOUT THE BUSINESS
- LOSS OF UNDERSTANDING MEANS LESS THAN SUCCESSFUL BUSINESS
  
- IT'S THE *SEMANTIC INFORMATION* THAT IS IMPORTANT TO THE BUSINESS
- THE MEANING OF THE DATA WITH RESPECT TO THE BUSINESS
- HOW WELL DO THE DATA REPRESENT THE BUSINESS

***LOW QUALITY DATA MEANS POOR BUSINESS UNDERSTANDING***



## 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!***



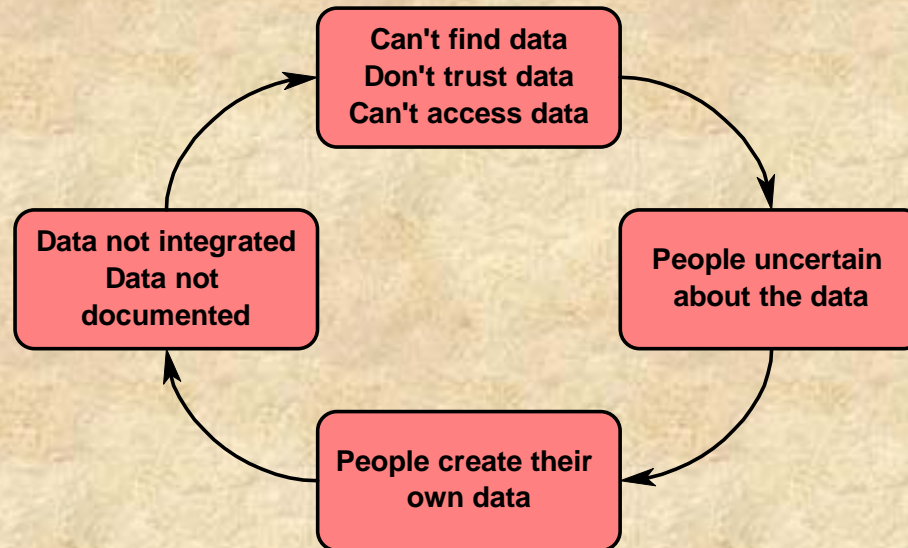
## CURRENT SITUATION

### ◆ **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!**



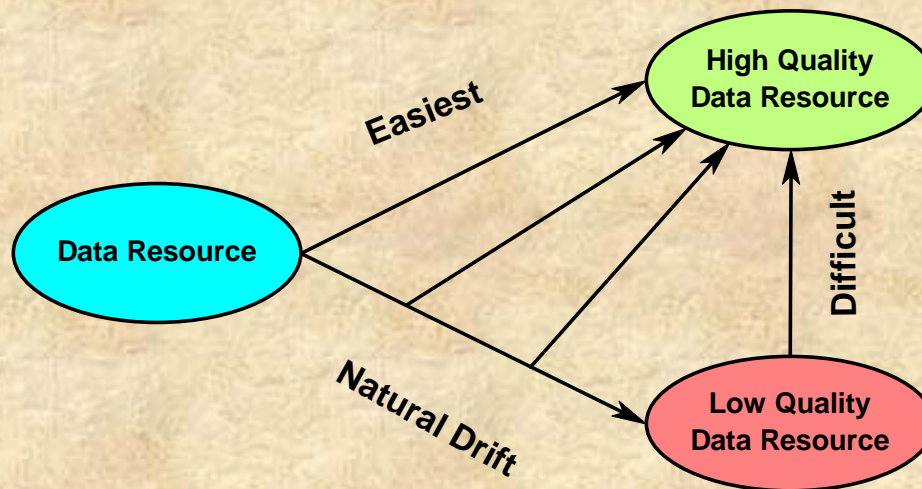
## CURRENT SITUATION

### ◆ **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

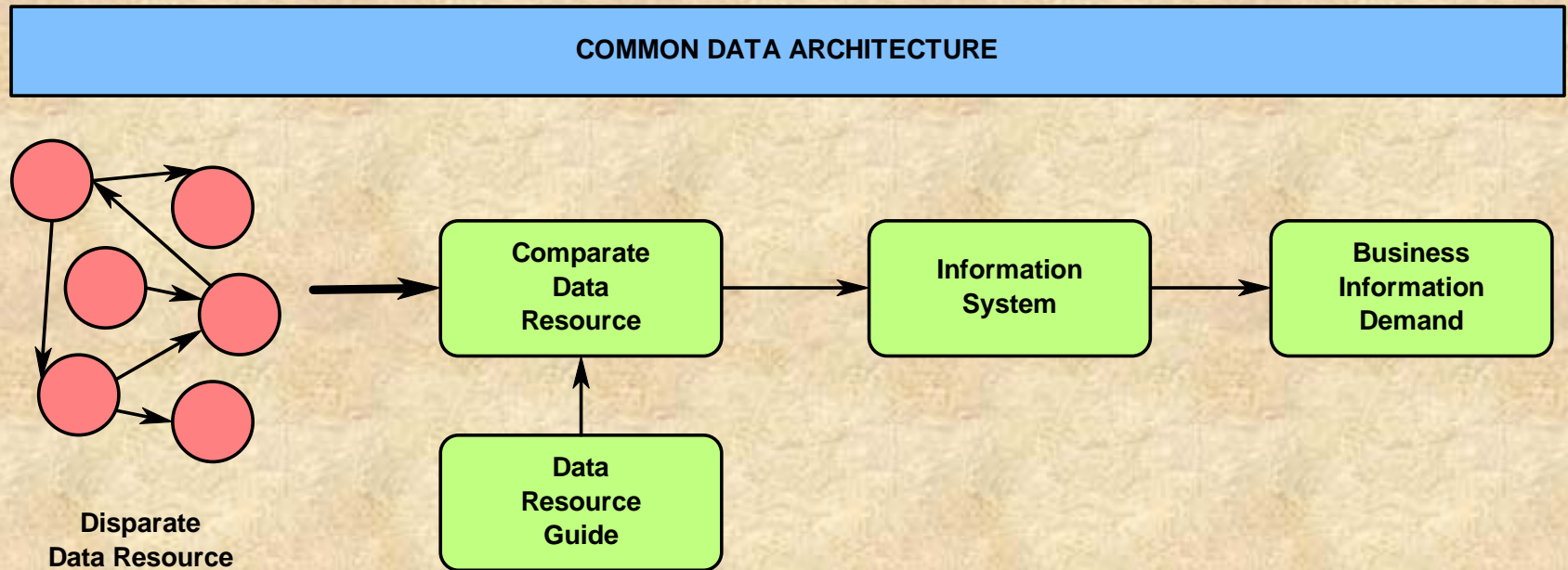
## ◆ COMPARE DATA VISION

- 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*

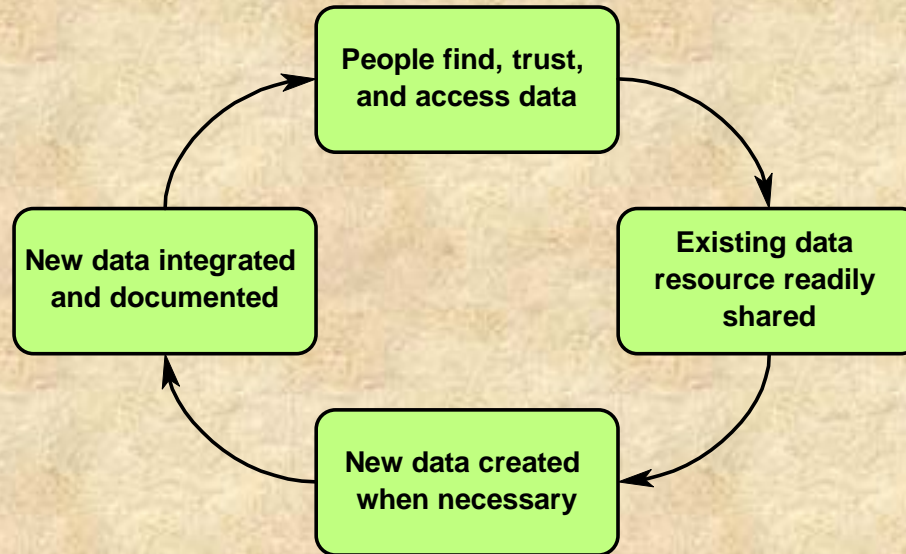




# DATA RESOURCE CONCEPTS

## ◆ **COMPARATE 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 COMPARATE DATA CYCLE!**



# DATA RESOURCE CONCEPTS

## ◆ **COMPARE 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!***



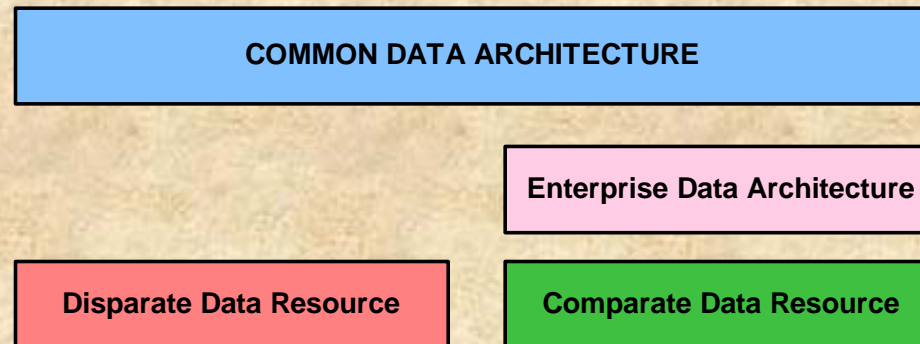
# DATA RESOURCE CONCEPTS

## ◆ 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

## ● DATA CHARACTERISTIC STRUCTURE

### Data Steward

Primary Key: Data Steward. System Identifier

Foreign Key: Data Steward Level

Data Steward Level. Name

Foreign Key: Organization Unit Identifier

Organization Unit. System

### Characteristic List:

Data Steward. Address City Name

Data Steward. Address Line 1

Data Steward. Address Line 2

Data Steward. Address Postal Code

Data Steward. Address State Postal Abbreviation

Data Steward. Begin Date

Data Steward. E-mail Address

Data Steward. End Date

Data Steward. Fax Number

Data Steward. Formal Name

Data Steward. Phone Number

Data Steward. System Identifier

Data Steward Level. Name

Organization Unit. System Identifier



# DATA RESOURCE CONCEPTS

- **DATA DEFINITIONS**

**Data Steward**

A Data Steward is a Person that is responsible for maintaining some aspect of an organization's data resource for the welfare of that organizations. A Data Steward can be responsible for existing Data Products or for Data Subjects in the Common Data Architecture.

**Data Steward. Address City Name**

The city name where the Organization Contact is located, such as 'Little Rock'.

**Data Steward. Address Line 1**

The first line of the address where the Data Steward is located according to the rules specified by the postal service.

**Data Steward. Address Line 2**

The second line of the address where the Data Steward is located according to the rules specified by the postal service.

**Data Steward. Address Postal Code**

The postal code for the Data Steward, such as '98765'.

**Data Steward. Address State Postal Abbreviation**

The Post Office abbreviation for the State Name, such as 'WA' for Washington.

**Data Steward. Begin Date**

The business date that the Data Steward became effective, such as 'January 15, 1987'. It is the first day that the Data Steward is effective.



# DATA RESOURCE CONCEPTS

- **DATA INTEGRITY RULES**

## **Data Steward**

Occurrence Deletes: Prevented

Proactive Updates: Allowed

Retroactive Updates: Prevented

### **Data Steward. Address City Name**

Need: Required

Domain: 5 characters <= text <= 40 characters & first character <> Blank

Value Change: Allowed with No History

### **Data Steward. Address Line 1**

Need: Required

Domain: 5 characters <= text <= 60 characters & first character <> Blank

Value Change: Allowed with No History

### **Data Steward. Address Line 2**

Need: Optional

Domain: {Blank | 5 characters <= text <= 60 characters & first character <> Blank}

Value Change: Allowed with No History

### **Data Steward. Address Postal Code**

Need: Required

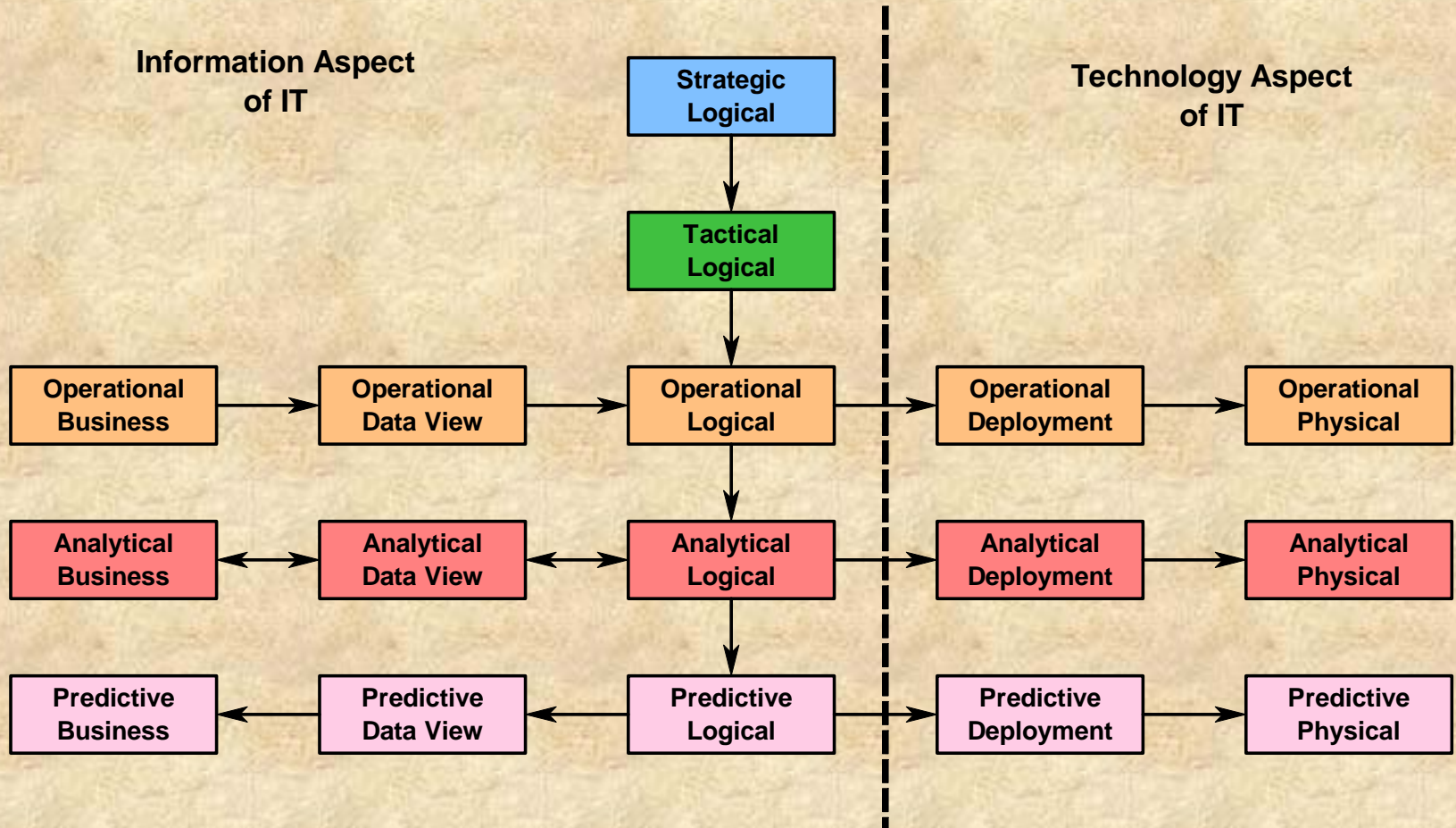
Domain: 5 characters <= text <= 10 characters & first character <> Blank

Value Change: Allowed with No History



# DATA RESOURCE CONCEPTS

## ◆ FIVE-TIER FIVE-SCHEMA CONCEPT

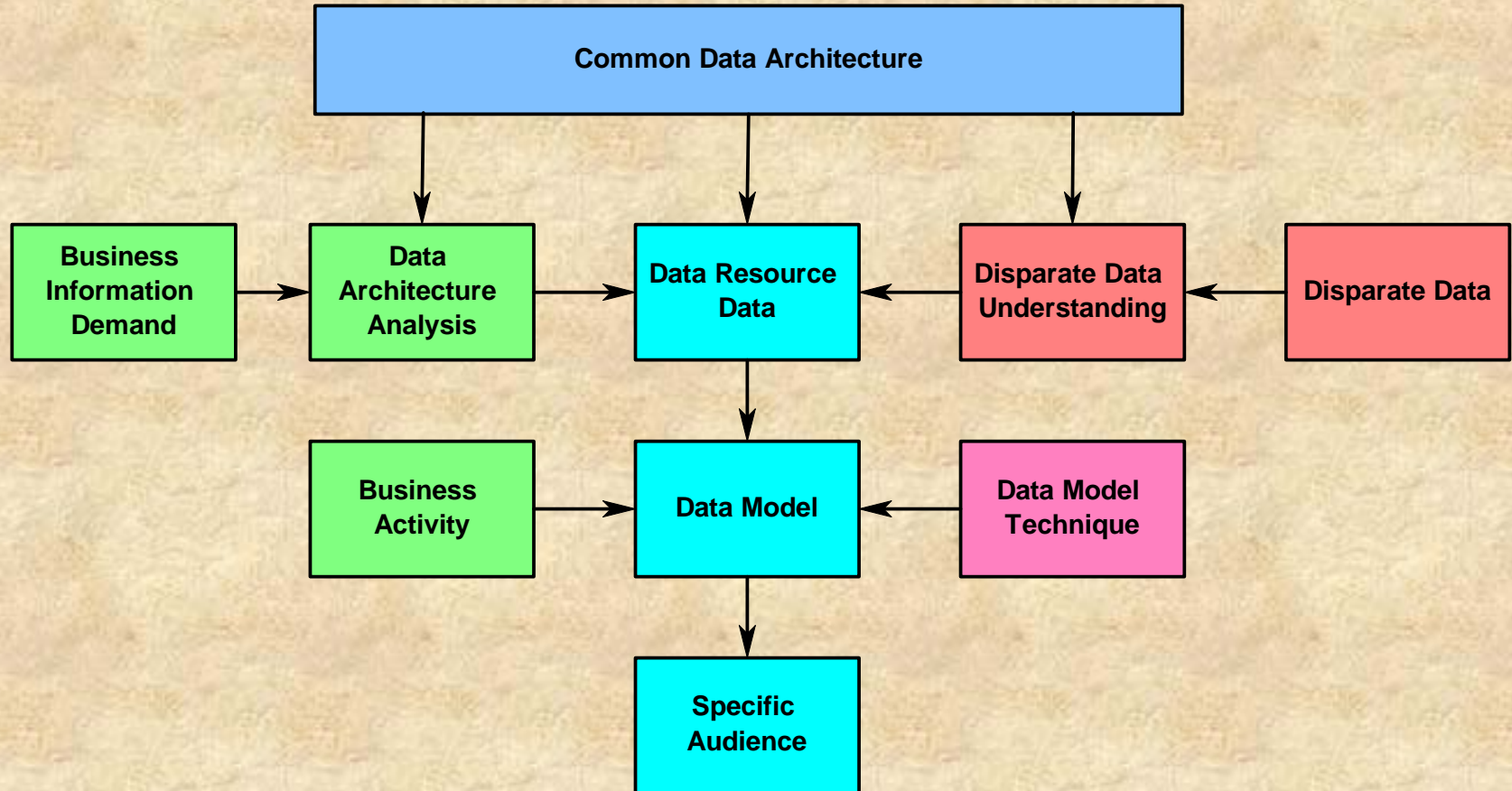


**CONCENTRATE ON INFORMATION ASPECT OF IT  
RATHER THAN TECHNOLOGY ASPECT!**



# DATA RESOURCE CONCEPTS

## ◆ ARCHITECTURE DRIVEN DATA MODELING



**DATA MODEL IS A REPRESENTATION OF A SUBSET OF THE DATA ARCHITECTURE FOR A SPECIFIC AUDIENCE!**



# DATA RESOURCE QUALITY

## ◆ **SEGMENTS OF QUALITY**

- ALL GOOD PRACTICES HAVE AN ENERGY AND A COST
- INDEPENDENT OF THE PRIMARY FOCUS
- TAKING A DIFFERENT FOCUS DOES NOT REDUCE THE COST

## ◆ **SILVER BULLETS**

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

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

- QUALITY WILL NOT HAPPEN BY ITSELF
- THE DATA RESOURCE WILL NOT SELF-ORGANIZE ITSELF
- THERE IS NO SELF-UNDERSTANDING IN THE DATA RESOURCE
- THERE IS NO INTERNAL SOURCE OF ENERGY
- EXTERNAL ENERGY REQUIRED FOR UNDERSTANDING AND REORGANIZATION
- THEREFORE THERE ARE NO SILVER BULLETS... AND NEVER WILL BE

***THERE ARE NO SILVER BULLETS***



# DATA RESOURCE QUALITY

## ◆ **COST EXAMPLES**

- **CAN BEGIN TO PUT VALUES ON COST OF QUALITY**
  
- **REDUCING REDUNDANCY TWO LEVELS = \$50 MILLION PER BIENNIUM**
- **PROPORTIONAL DOWN TO A REDUNDANCY FACTOR OF 4**
  
- **INVENTORIED 250,000 DATA ITEMS**
- **PRESUMED REDUNDANCY FACTOR OF 10**
- **EQUALS 25,000 ATTRIBUTES AND 2,000 ENTITIES**
- **BUSINESS ANALYSIS SHOWED 1,200 ENTITIES AND 12,000 ATTRIBUTES**
- **DON'T KNOW THE BUSINESS OR REDUNDANCY FACTOR OF 20**
  
- **HEIGHT IN OUNCES**
- **TOOK 2 PEOPLE TOTAL OF 4 HOURS TO RESOLVE**
- **EXTRAPOLATED TO \$12 MILLION PER YEAR TO UNDERSTAND DATA**
  
- **23 CUSTOMER FILES**
- **STILL CALCULATING THE COST OF NOT KNOWING CUSTOMER BASE**

***LOW QUALITY IS VERY EXPENSIVE***





# DATA RESOURCE QUALITY

## ◆ ACCEPTABLE LEVEL OF QUALITY

- QUALITY – THE REDUCTION OF ENTROPY – IS INFINITE
- THE COST COULD BE INFINITE FOR ABSOLUTE PERFECTION
  
- BUSINESS MUST DETERMINE THE ACCEPTABLE LEVEL OF QUALITY
- WHAT LEVEL OF QUALITY IS APPROPRIATE FOR THE BUSINESS
- WHAT LEVEL OF ENERGY / COST IS APPROPRIATE TO MAINTAIN QUALITY
- WHAT LEVEL OF ENERGY / COST IS APPROPRIATE TO IMPROVE QUALITY
- BASED ON WHAT IT TAKES TO UNDERSTAND AND OPERATE THE BUSINESS
  
- HOW GOOD IS GOOD ENOUGH?
- MODIFY EINSTEIN'S PRINCIPLE
- DATA RESOURCE QUALITY SHOULD BE GOOD ENOUGH TO RUN THE BUSINESS SUCCESSFULLY - AND NO BETTER
- THEN BASE COST ON THAT LEVEL OF QUALITY DESIRED

***THE DATA RESOURCE CANNOT BE SCRAPPED***



# CONCLUSION

## ◆ FOUNDATION

- FOUNDATION FOR QUALITY BASED ON PHYSICS
- DATA RESOURCE IS AN OPEN SYSTEM
- ENTROPY INCREASES OVER TIME – QUALITY DECREASES
- NEED ENERGY TO MAINTAIN AND IMPROVE DATA RESOURCE QUALITY
- NEED QUALITY DATA TO UNDERSTAND THE BUSINESS

## ◆ DATA RESOURCE

- DEFINED DATA AND INFORMATION QUALITY
- TEN GOOD PRACTICES FOR DATA RESOURCE QUALITY
- COST OF POOR QUALITY
- ACCEPTABLE LEVEL OF QUALITY
- COST OF QUALITY BASED ON SIZE – HISTORY – BUSINESS NEEDS
- NEW CONCEPTS FOR QUALITY

## ◆ SUMMARY

- QUALITY IS NOT FREE – BUT IT'S LESS EXPENSIVE THAN THE IMPACTS
- DATA RESOURCE CANNOT BE SCRAPPED – IT MUST BE IMPROVED
- IMPROVEMENT SUFFICIENT TO MEET BUSINESS NEEDS



## 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**