The MIT 2008 Information Quality Industry Symposium

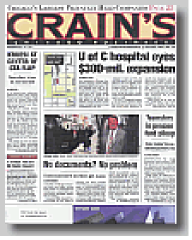
Using Conceptual Data Modeling to ensure high Information and Data Quality

Pete Stiglich
Senior Consultant
PStiglich@ewsolutions.com

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 1
Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

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 clients a full range of services for any size enterprise information management, managed meta data environment, and/or data warehouse/business intelligence initiative. Our notable client projects have been featured in the Chicago Tribune, Federal Computer Weekly, Crain's Chicago Business, and won the 2004 Intelligent Enterprise's RealWare award, 2007 Excellence in Information Integrity Award nomination and DM Review's 2005 World Class Solutions award.



Information Integrity Coalition

2007 Excellence in Information Integrity Award Nomination

Chicago Tribune

DM Review 2005

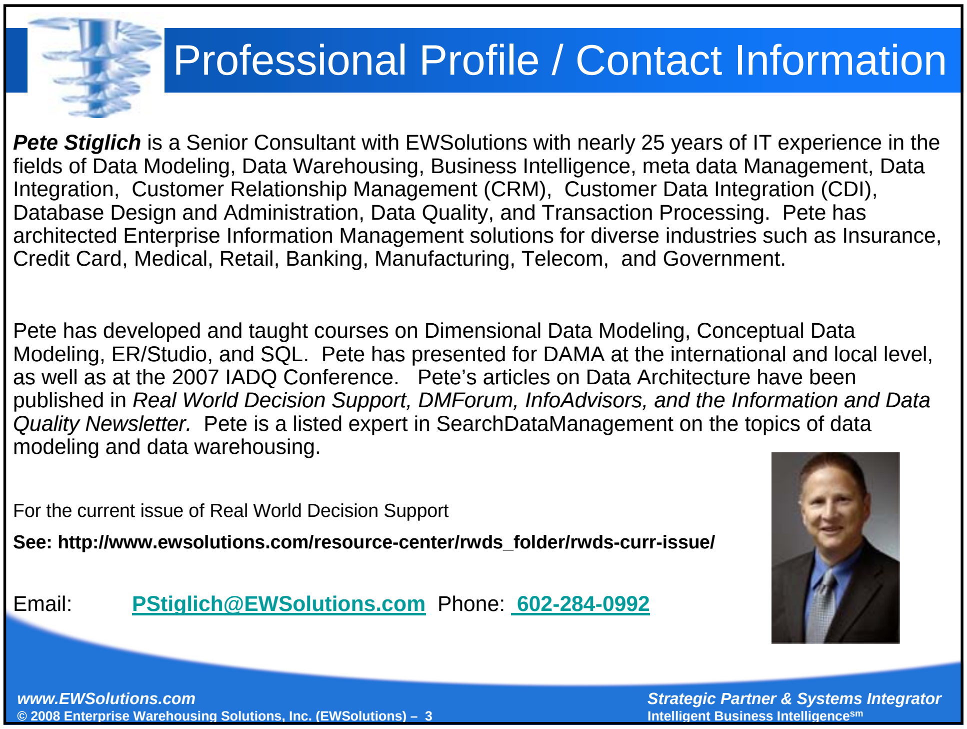
World Class Solutions Award Data Management

intelligent enterprise REAL WARE Awards TRANSFORM 2004 WINNER

Best Business Intelligence Application
Information Integration
Client: Department of Defense

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, call toll free at 866.EWS.1100, 866.397.1100, mai number 630.920.0005 or email us at Info@EWSolutions.com

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 2
Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



Professional Profile / Contact Information

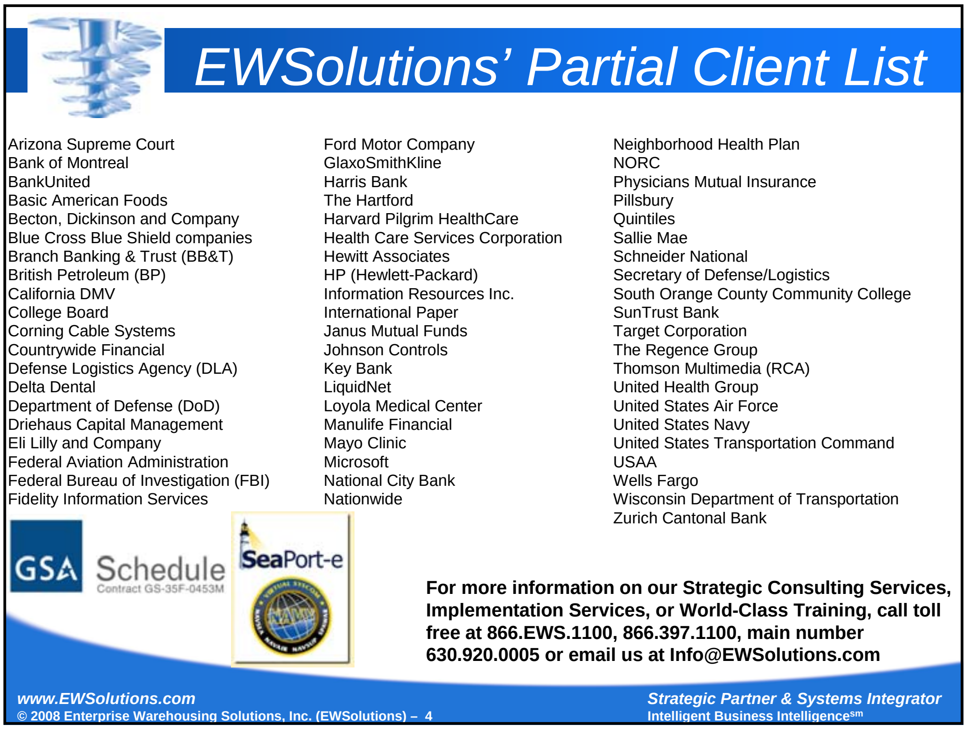
Pete Stiglich is a Senior Consultant with EWSolutions with nearly 25 years of IT experience in the fields of Data Modeling, Data Warehousing, Business Intelligence, meta data Management, Data Integration, Customer Relationship Management (CRM), Customer Data Integration (CDI), Database Design and Administration, Data Quality, and Transaction Processing. Pete has architected Enterprise Information Management solutions for diverse industries such as Insurance, Credit Card, Medical, Retail, Banking, Manufacturing, Telecom, and Government.

Pete has developed and taught courses on Dimensional Data Modeling, Conceptual Data Modeling, ER/Studio, and SQL. Pete has presented for DAMA at the international and local level, as well as at the 2007 IADQ Conference. Pete's articles on Data Architecture have been published in *Real World Decision Support, DMForum, InfoAdvisors, and the Information and Data Quality Newsletter.* Pete is a listed expert in SearchDataManagement on the topics of data modeling and data warehousing.

For the current issue of Real World Decision Support

See: http://www.ewsolutions.com/resource-center/rwds_folder/rwds-curr-issue/

Email: PStiglich@EWSolutions.com Phone: 602-284-0992



EWSolutions' Partial Client List

Arizona Supreme Court
Bank of Montreal
BankUnited
Basic American Foods
Becton, Dickinson and Company
Blue Cross Blue Shield companies
Branch Banking & Trust (BB&T)
British Petroleum (BP)
California DMV
College Board
Corning Cable Systems
Countrywide Financial
Defense Logistics Agency (DLA)
Delta Dental
Department of Defense (DoD)
Driehaus Capital Management
Eli Lilly and Company
Federal Aviation Administration
Federal Bureau of Investigation (FBI)
Fidelity Information Services

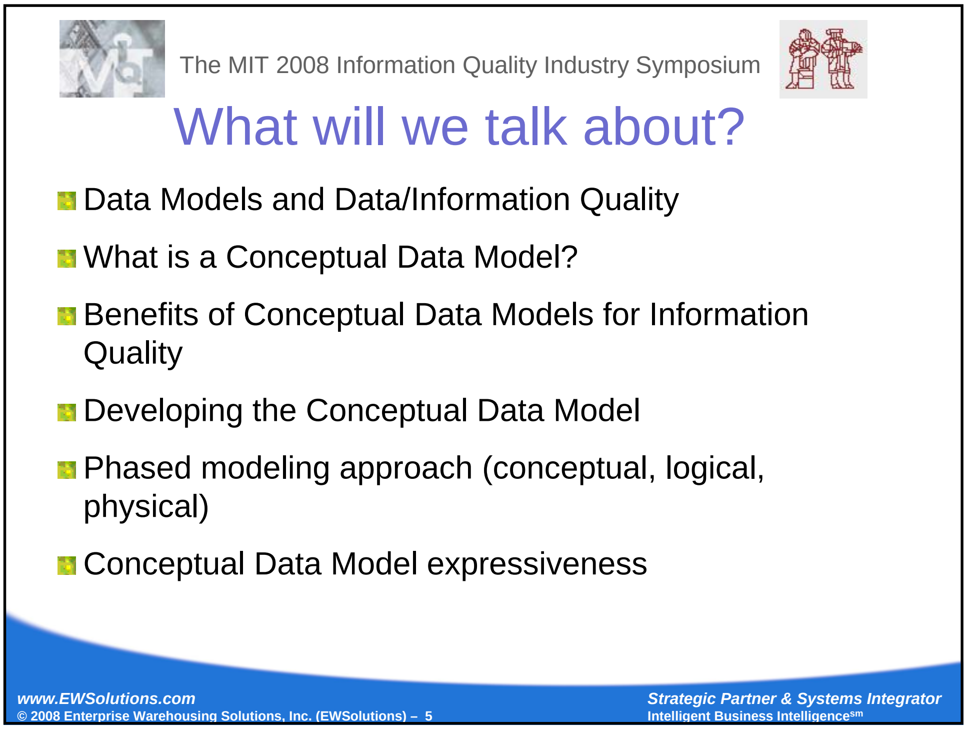
Ford Motor Company
GlaxoSmithKline
Harris Bank
The Hartford
Harvard Pilgrim HealthCare
Health Care Services Corporation
Hewitt Associates
HP (Hewlett-Packard)
Information Resources Inc.
International Paper
Janus Mutual Funds
Johnson Controls
Key Bank
LiquidNet
Loyola Medical Center
Manulife Financial
Mayo Clinic
Microsoft
National City Bank
Nationwide

Neighborhood Health Plan
NORC
Physicians Mutual Insurance
Pillsbury
Quintiles
Sallie Mae
Schneider National
Secretary of Defense/Logistics
South Orange County Community College
SunTrust Bank
Target Corporation
The Regence Group
Thomson Multimedia (RCA)
United Health Group
United States Air Force
United States Navy
United States Transportation Command
USAA
Wells Fargo
Wisconsin Department of Transportation
Zurich Cantonal Bank

GSA Schedule
Contract GS-35F-0453M

SeaPort-e

For more information on our Strategic Consulting Services, Implementation Services, or World-Class Training, call toll free at 866.EWS.1100, 866.397.1100, main number 630.920.0005 or email us at Info@EWSolutions.com



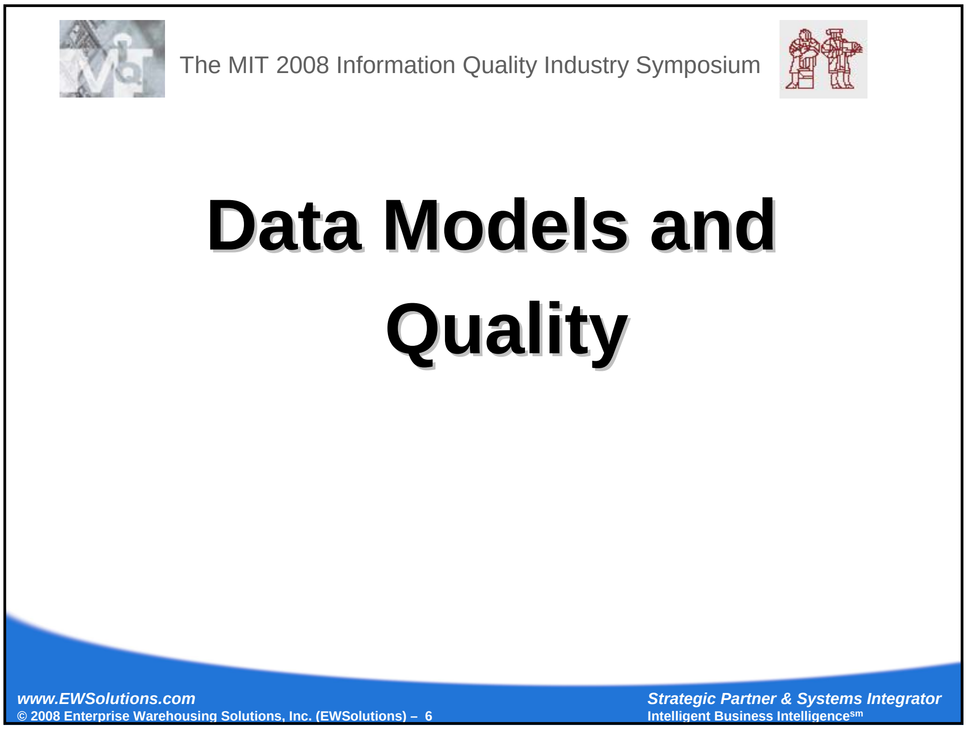
The MIT 2008 Information Quality Industry Symposium

What will we talk about?

- Data Models and Data/Information Quality
- What is a Conceptual Data Model?
- Benefits of Conceptual Data Models for Information Quality
- Developing the Conceptual Data Model
- Phased modeling approach (conceptual, logical, physical)
- Conceptual Data Model expressiveness

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 5

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

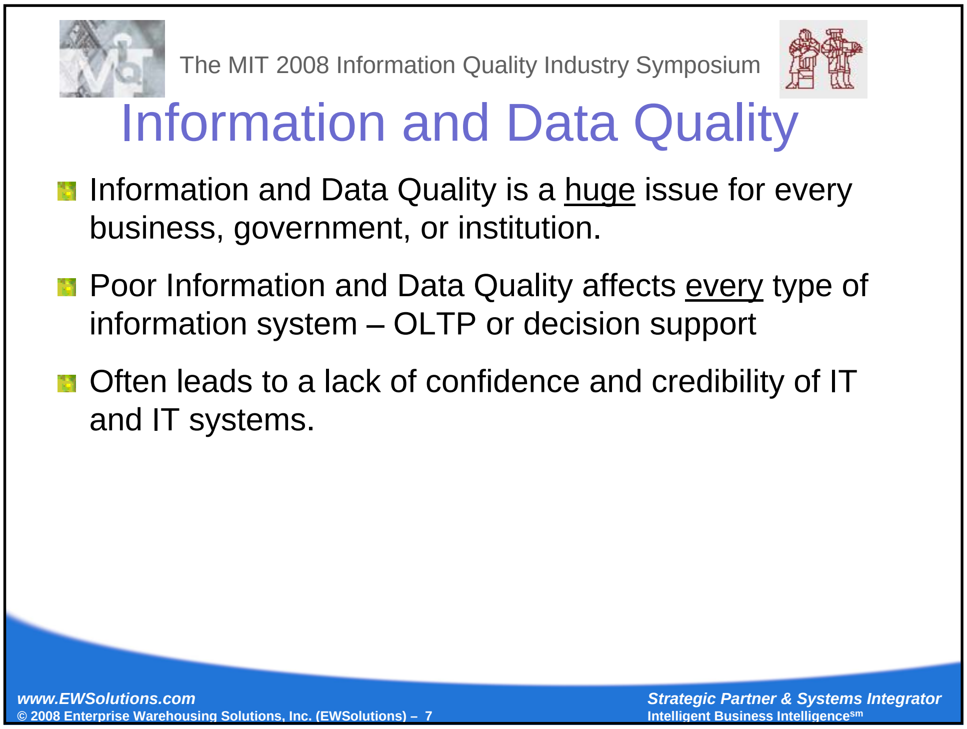


The MIT 2008 Information Quality Industry Symposium

Data Models and Quality

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 6

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Information and Data Quality

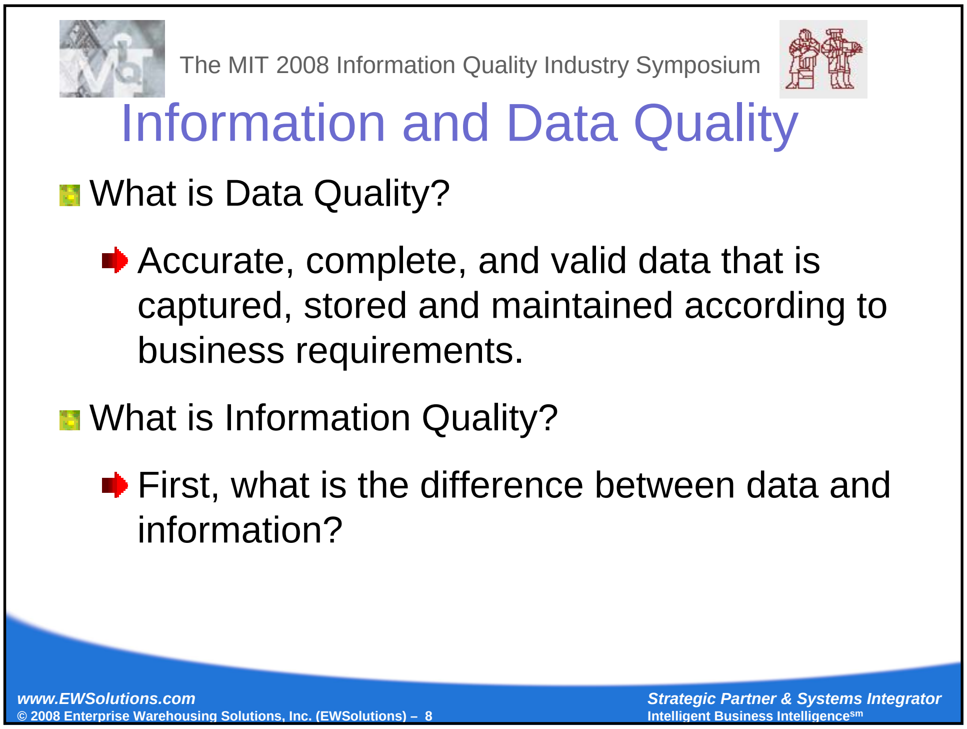
- Information and Data Quality is a huge issue for every business, government, or institution.
- Poor Information and Data Quality affects every type of information system – OLTP or decision support
- Often leads to a lack of confidence and credibility of IT and IT systems.

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 7

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

Information and Data Quality

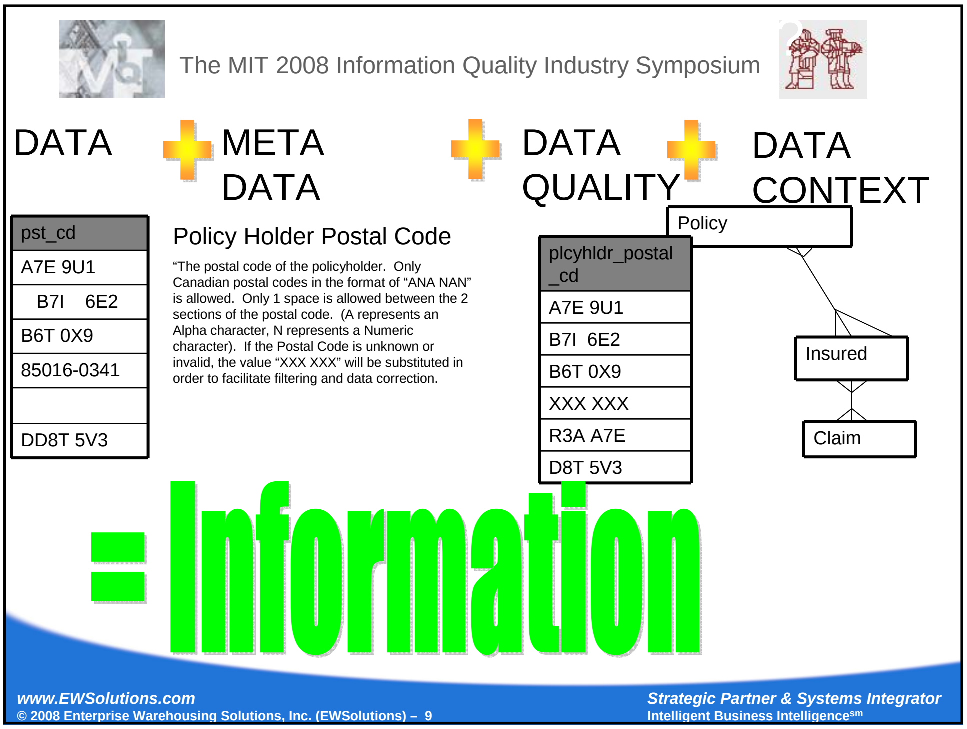
- What is Data Quality?
  - Accurate, complete, and valid data that is captured, stored and maintained according to business requirements.
- What is Information Quality?
  - First, what is the difference between data and information?

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 8

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

DATA + META DATA + DATA QUALITY + DATA CONTEXT

pst_cd
A7E 9U1
B7I 6E2
B6T 0X9
85016-0341
DD8T 5V3

Policy Holder Postal Code

"The postal code of the policyholder. Only Canadian postal codes in the format of "ANA NAN" is allowed. Only 1 space is allowed between the 2 sections of the postal code. (A represents an Alpha character, N represents a Numeric character). If the Postal Code is unknown or invalid, the value "XXX XXX" will be substituted in order to facilitate filtering and data correction.

plcyhldr_postal_cd
A7E 9U1
B7I 6E2
B6T 0X9
XXX XXX
R3A A7E
D8T 5V3

Policy

Insured

Claim

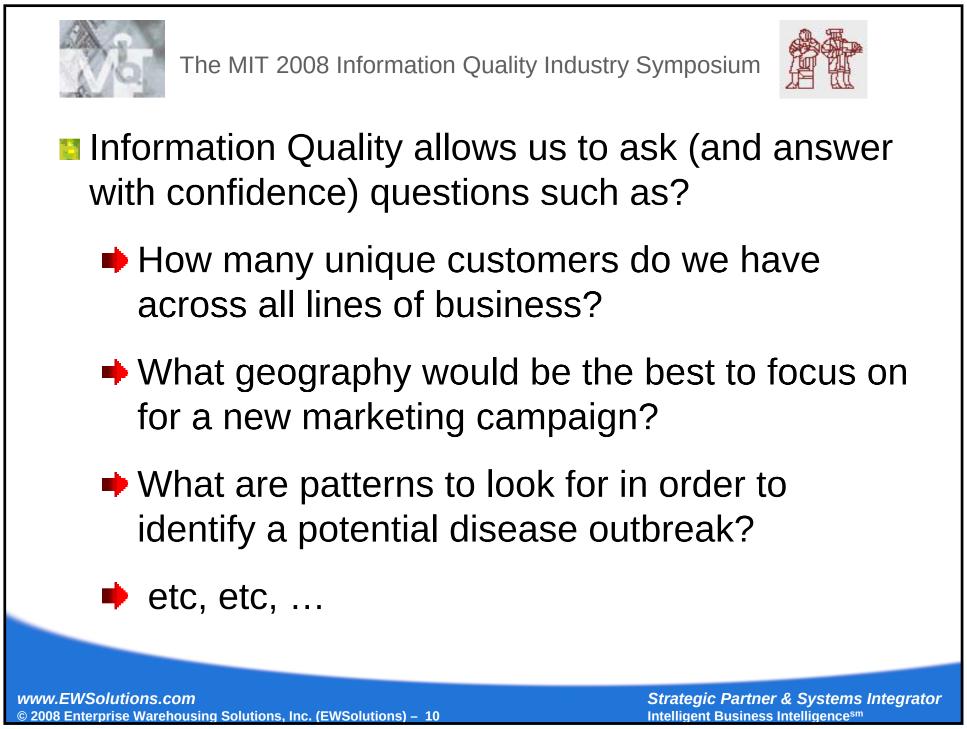
= Information

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 9

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

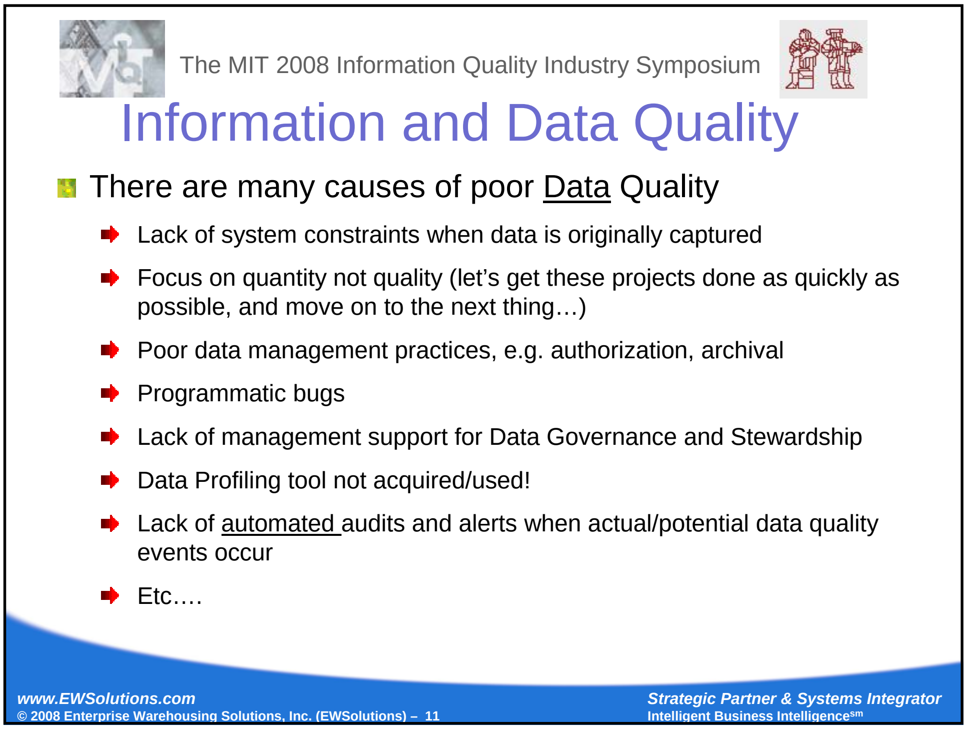
- Information Quality allows us to ask (and answer with confidence) questions such as?
  - How many unique customers do we have across all lines of business?
  - What geography would be the best to focus on for a new marketing campaign?
  - What are patterns to look for in order to identify a potential disease outbreak?
  - etc, etc, …

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 10

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Information and Data Quality

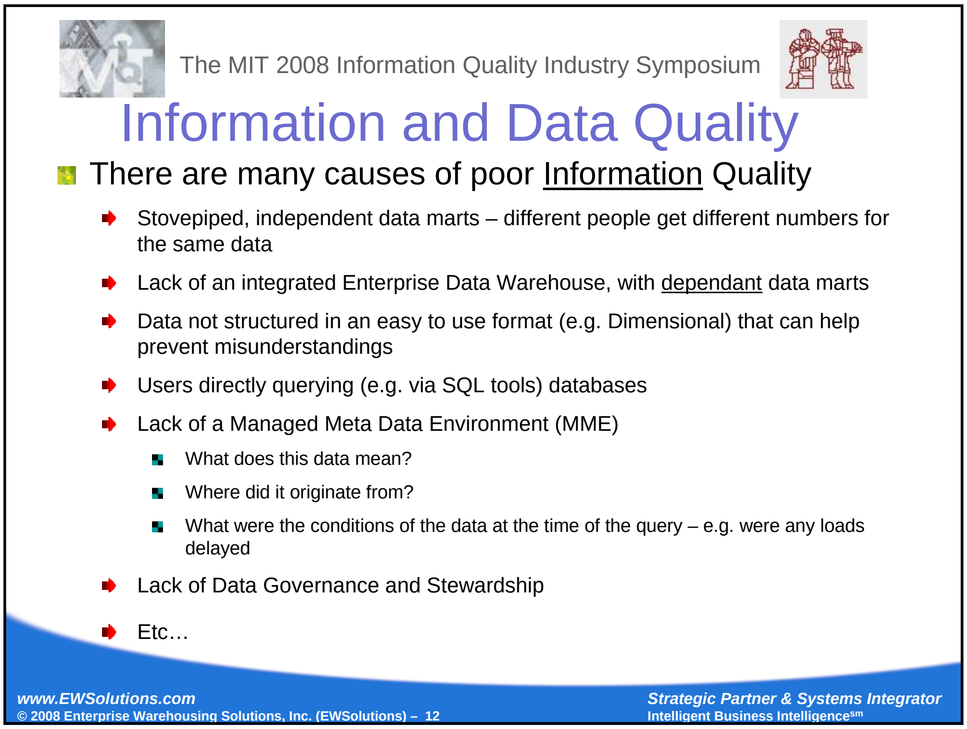
- There are many causes of poor Data Quality
  - Lack of system constraints when data is originally captured
  - Focus on quantity not quality (let's get these projects done as quickly as possible, and move on to the next thing…)
  - Poor data management practices, e.g. authorization, archival
  - Programmatic bugs
  - Lack of management support for Data Governance and Stewardship
  - Data Profiling tool not acquired/used!
  - Lack of automated audits and alerts when actual/potential data quality events occur
  - Etc.…

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 11

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Information and Data Quality

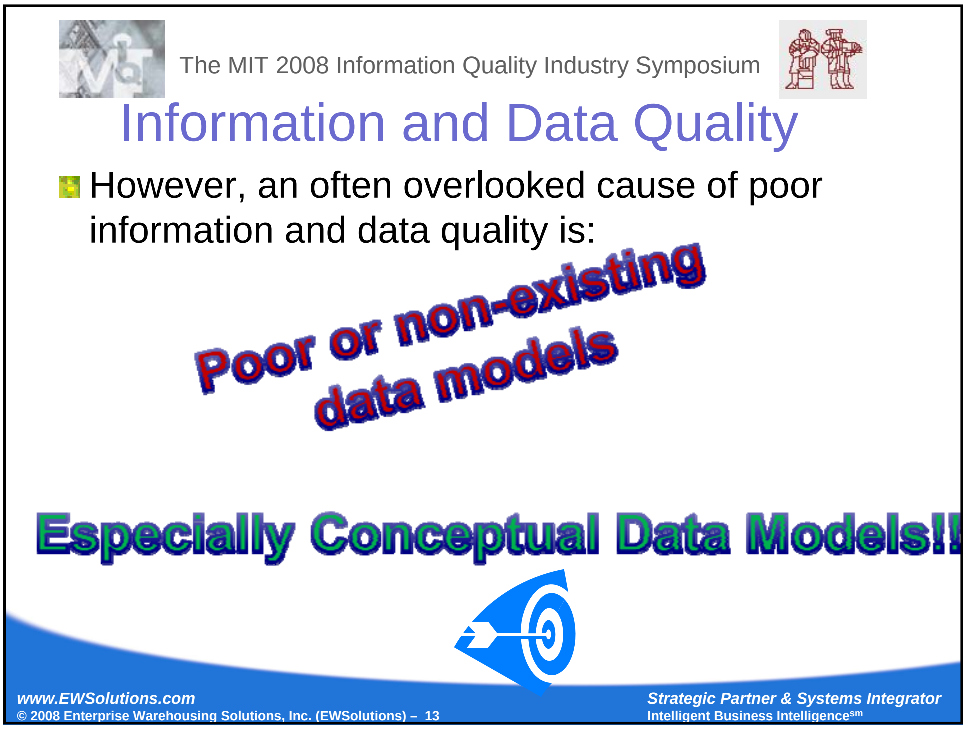
- There are many causes of poor Information Quality
  - Stovepiped, independent data marts – different people get different numbers for the same data
  - Lack of an integrated Enterprise Data Warehouse, with dependant data marts
  - Data not structured in an easy to use format (e.g. Dimensional) that can help prevent misunderstandings
  - Users directly querying (e.g. via SQL tools) databases
  - Lack of a Managed Meta Data Environment (MME)
    - What does this data mean?
    - Where did it originate from?
    - What were the conditions of the data at the time of the query – e.g. were any loads delayed
  - Lack of Data Governance and Stewardship
  - Etc…

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 12

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Information and Data Quality

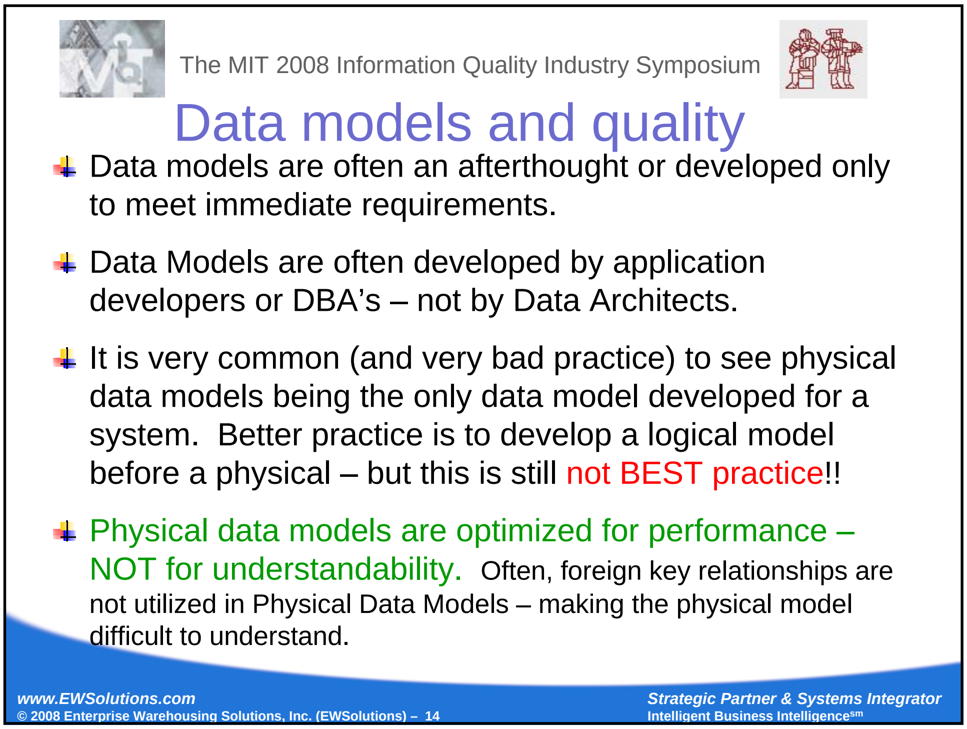
- However, an often overlooked cause of poor information and data quality is:

Poor or non-existing data models

Especially Conceptual Data Models!!

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 13

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



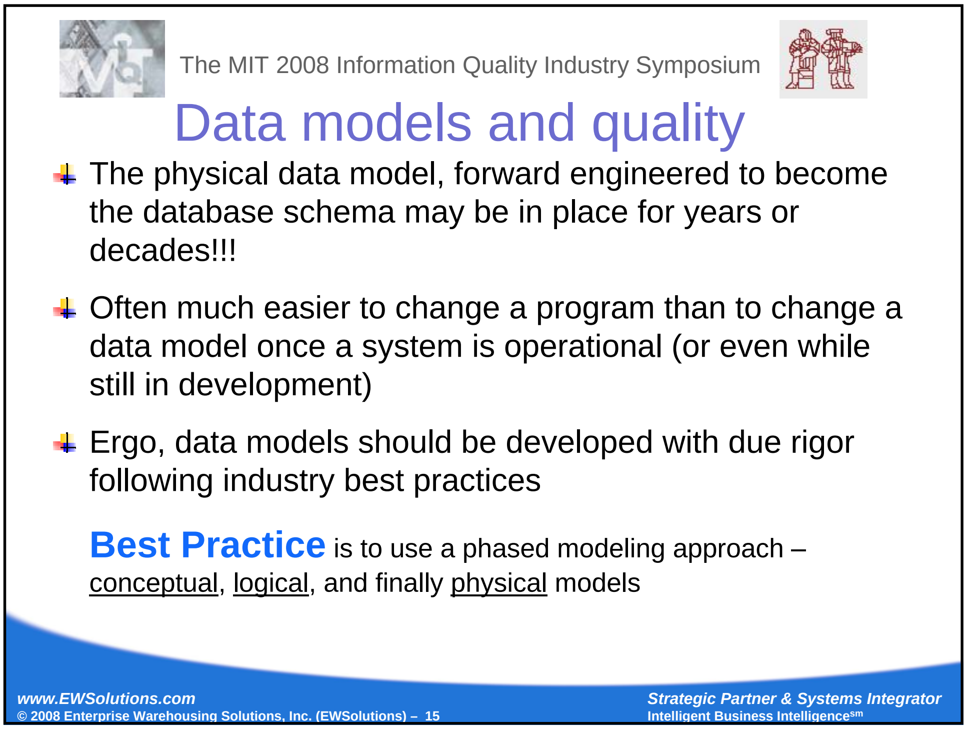
The MIT 2008 Information Quality Industry Symposium

Data models and quality

- Data models are often an afterthought or developed only to meet immediate requirements.
- Data Models are often developed by application developers or DBA’s – not by Data Architects.
- It is very common (and very bad practice) to see physical data models being the only data model developed for a system. Better practice is to develop a logical model before a physical – but this is still not BEST practice!!
- Physical data models are optimized for performance – NOT for understandability. Often, foreign key relationships are not utilized in Physical Data Models – making the physical model difficult to understand.

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 14

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

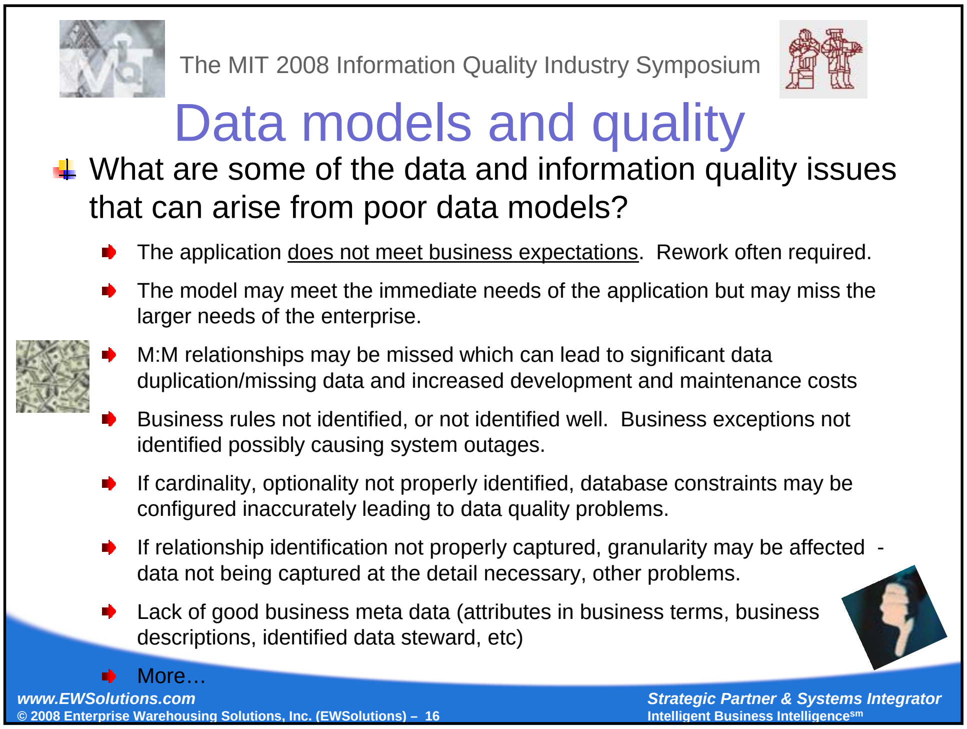
Data models and quality

- The physical data model, forward engineered to become the database schema may be in place for years or decades!!!
- Often much easier to change a program than to change a data model once a system is operational (or even while still in development)
- Ergo, data models should be developed with due rigor following industry best practices

Best Practice is to use a phased modeling approach – conceptual, logical, and finally physical models

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 15

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



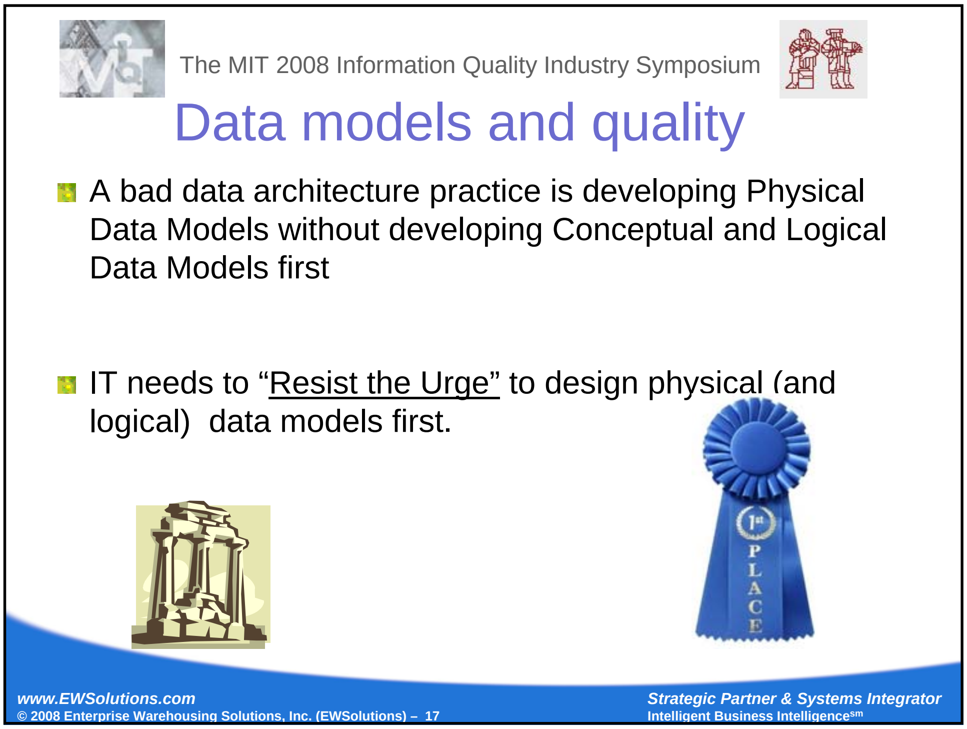
The MIT 2008 Information Quality Industry Symposium

Data models and quality

- What are some of the data and information quality issues that can arise from poor data models?
  - The application does not meet business expectations. Rework often required.
  - The model may meet the immediate needs of the application but may miss the larger needs of the enterprise.
  - M:M relationships may be missed which can lead to significant data duplication/missing data and increased development and maintenance costs
  - Business rules not identified, or not identified well. Business exceptions not identified possibly causing system outages.
  - If cardinality, optionality not properly identified, database constraints may be configured inaccurately leading to data quality problems.
  - If relationship identification not properly captured, granularity may be affected - data not being captured at the detail necessary, other problems.
  - Lack of good business meta data (attributes in business terms, business descriptions, identified data steward, etc)
  - More…

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 16

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



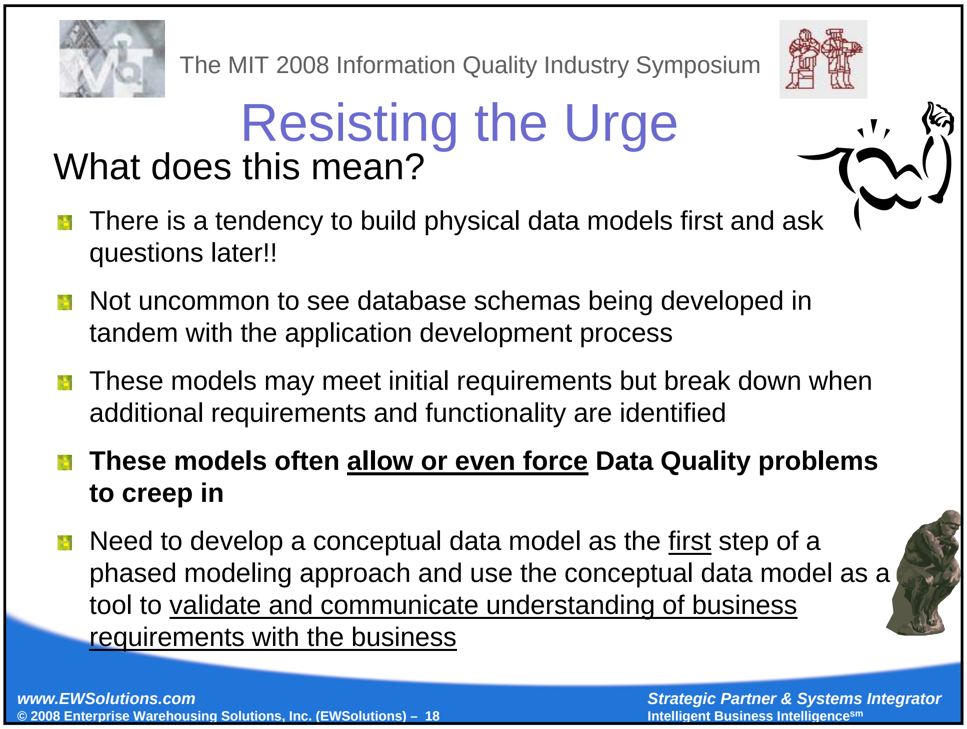
The MIT 2008 Information Quality Industry Symposium

Data models and quality

- A bad data architecture practice is developing Physical Data Models without developing Conceptual and Logical Data Models first
- IT needs to "Resist the Urge" to design physical (and logical) data models first.

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 17

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

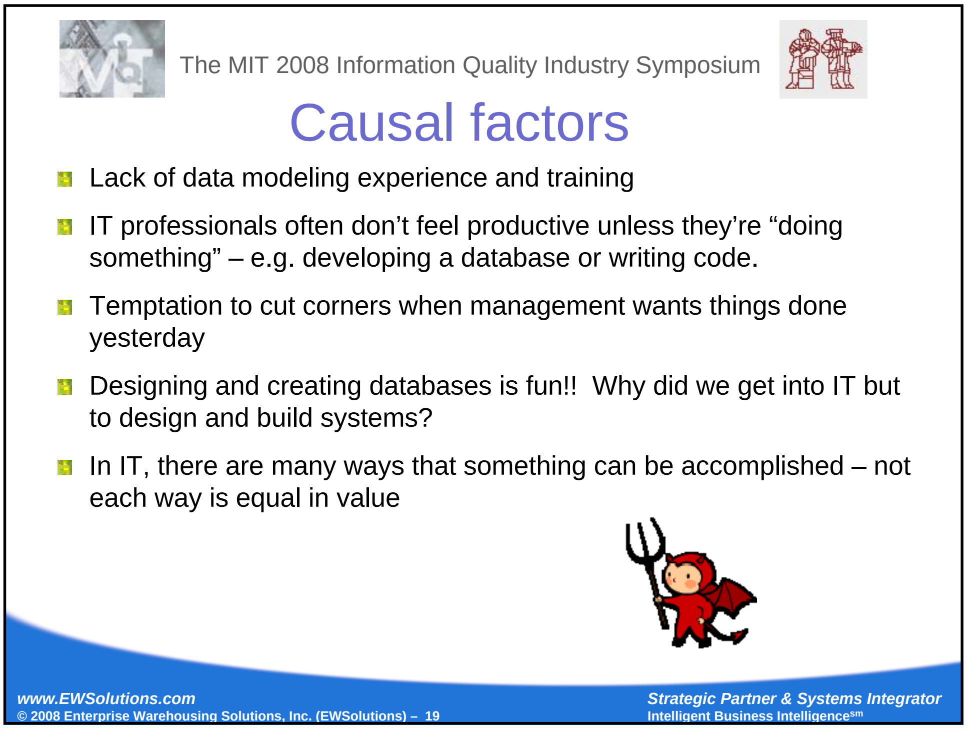
Resisting the Urge

What does this mean?

- There is a tendency to build physical data models first and ask questions later!!
- Not uncommon to see database schemas being developed in tandem with the application development process
- These models may meet initial requirements but break down when additional requirements and functionality are identified
- **These models often allow or even force Data Quality problems to creep in**
- Need to develop a conceptual data model as the first step of a phased modeling approach and use the conceptual data model as a tool to validate and communicate understanding of business requirements with the business

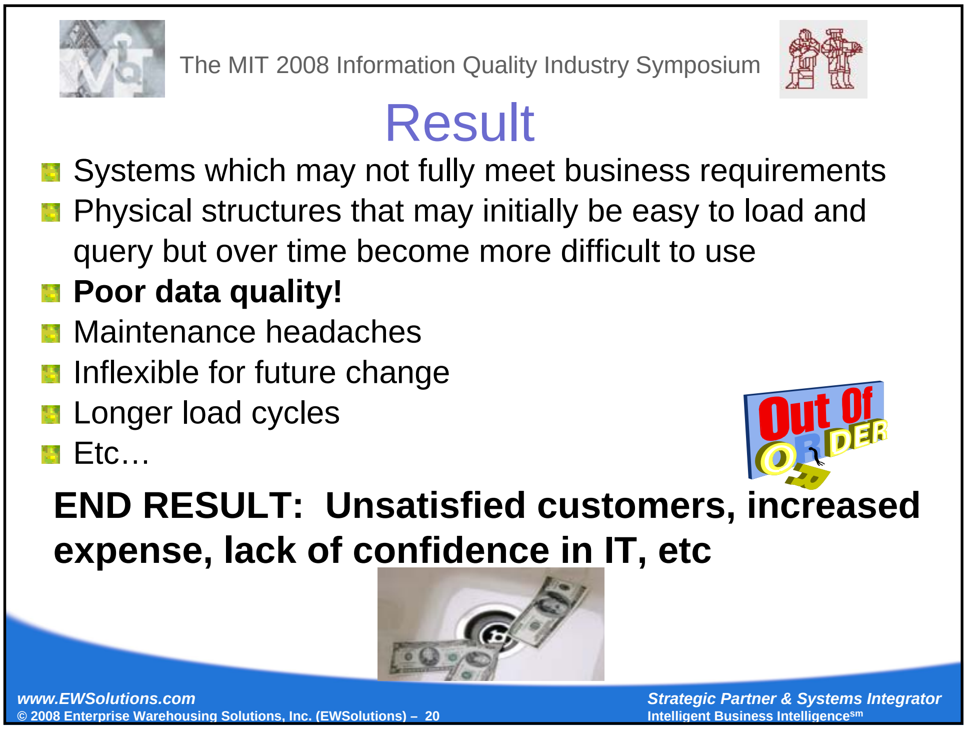
www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 18

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



Causal factors

- Lack of data modeling experience and training
- IT professionals often don’t feel productive unless they’re “doing something” – e.g. developing a database or writing code.
- Temptation to cut corners when management wants things done yesterday
- Designing and creating databases is fun!! Why did we get into IT but to design and build systems?
- In IT, there are many ways that something can be accomplished – not each way is equal in value



Result

- Systems which may not fully meet business requirements
- Physical structures that may initially be easy to load and query but over time become more difficult to use
- **Poor data quality!**
- Maintenance headaches
- Inflexible for future change
- Longer load cycles
- Etc…

Out Of
ORDER

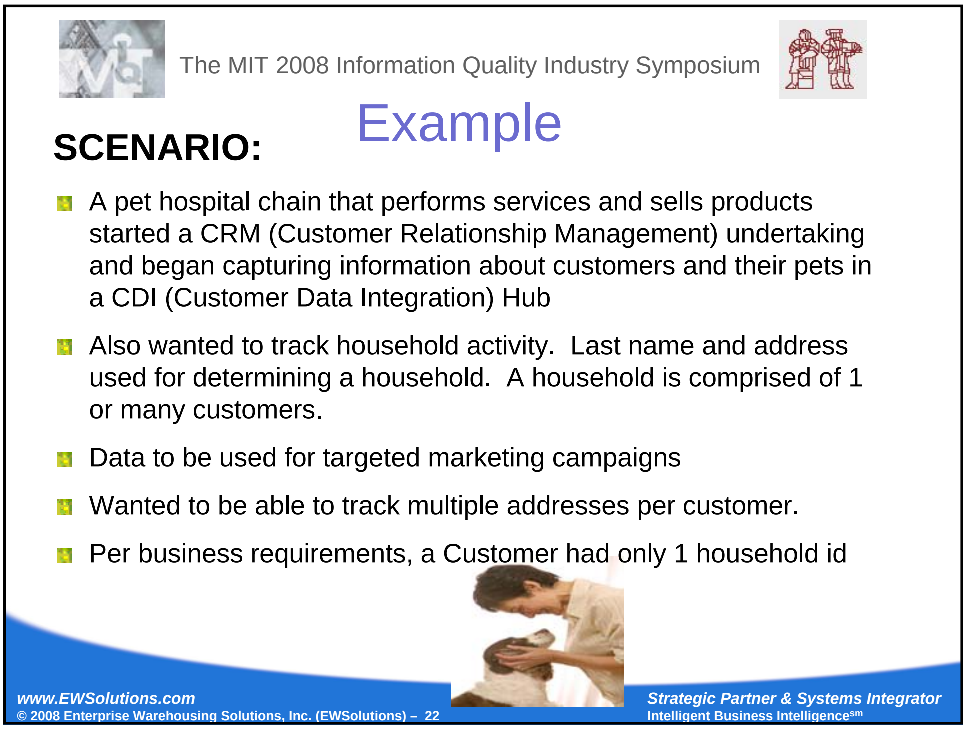
END RESULT: Unsatisfied customers, increased expense, lack of confidence in IT, etc

The MIT 2008 Information Quality Industry Symposium

Group Exercise

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 21

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

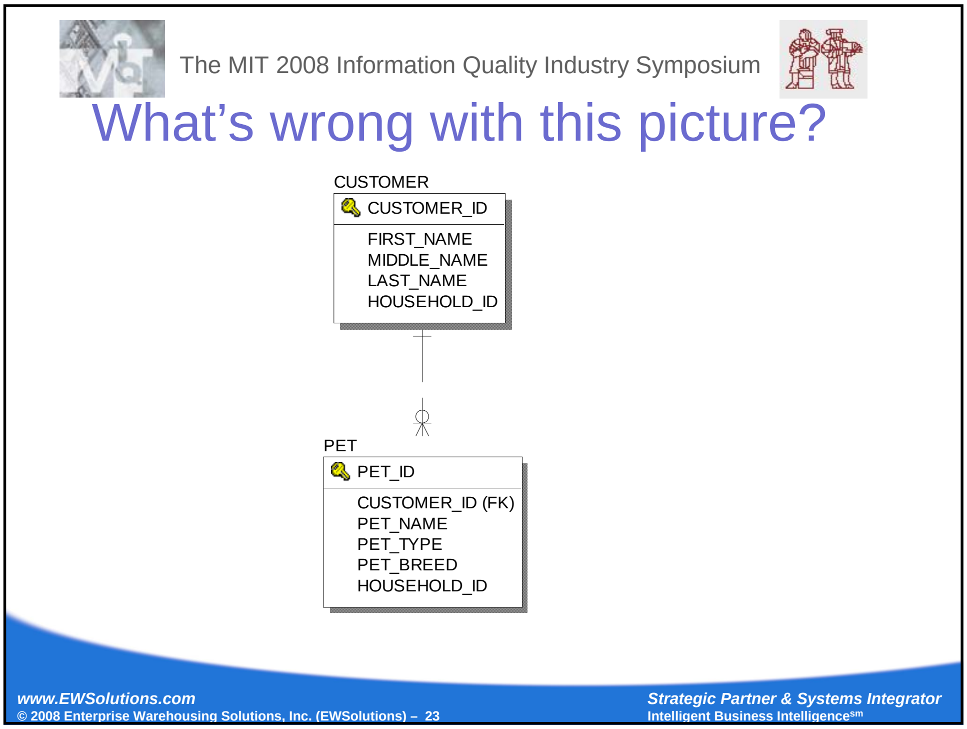
Example

SCENARIO:

- A pet hospital chain that performs services and sells products started a CRM (Customer Relationship Management) undertaking and began capturing information about customers and their pets in a CDI (Customer Data Integration) Hub
- Also wanted to track household activity. Last name and address used for determining a household. A household is comprised of 1 or many customers.
- Data to be used for targeted marketing campaigns
- Wanted to be able to track multiple addresses per customer.
- Per business requirements, a Customer had only 1 household id

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 22

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

What's wrong with this picture?

CUSTOMER

CUSTOMER_ID

FIRST_NAME
MIDDLE_NAME
LAST_NAME
HOUSEHOLD_ID

PET

PET_ID

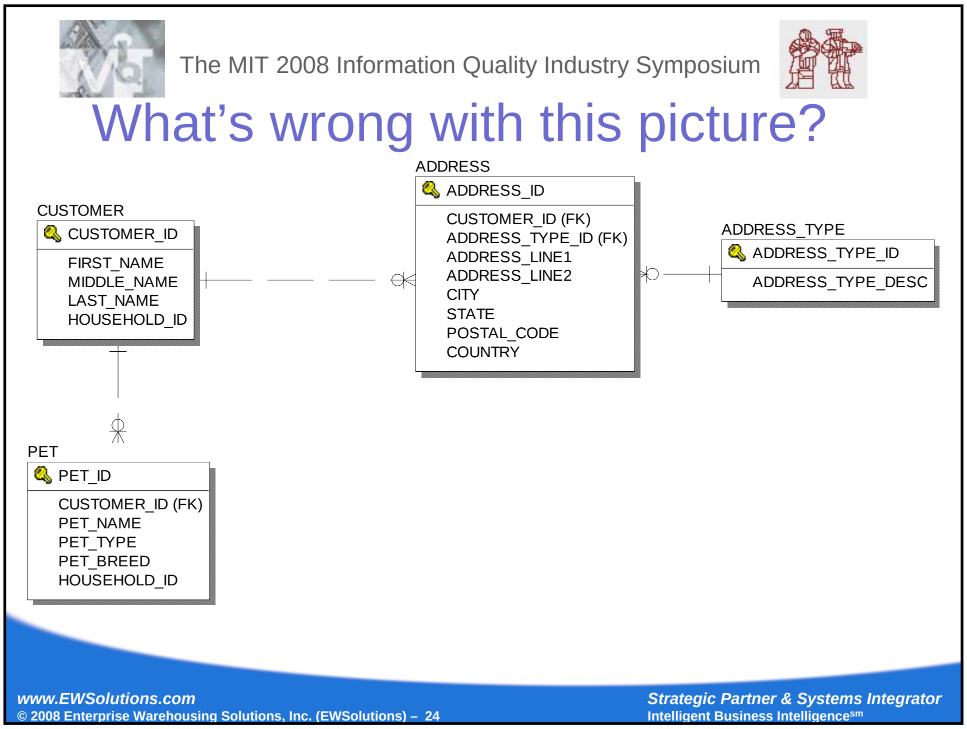
CUSTOMER_ID (FK)
PET_NAME
PET_TYPE
PET_BREED
HOUSEHOLD_ID

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 23

Strategic Partner & Systems Integrator

Intelligent Business Intelligencesm



The MIT 2008 Information Quality Industry Symposium

What's wrong with this picture?

ADDRESS

ADDRESS_ID

CUSTOMER_ID (FK)
ADDRESS_TYPE_ID (FK)
ADDRESS_LINE1
ADDRESS_LINE2
CITY
STATE
POSTAL_CODE
COUNTRY

CUSTOMER

CUSTOMER_ID

FIRST_NAME
MIDDLE_NAME
LAST_NAME
HOUSEHOLD_ID

ADDRESS_TYPE

ADDRESS_TYPE_ID

ADDRESS_TYPE_DESC

PET

PET_ID

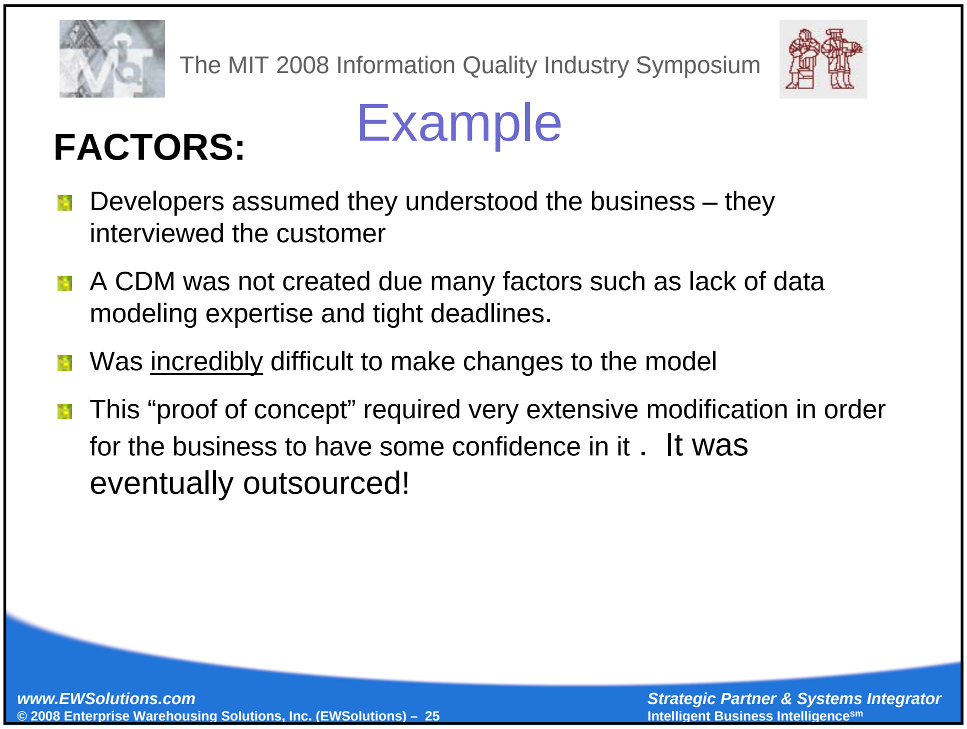
CUSTOMER_ID (FK)
PET_NAME
PET_TYPE
PET_BREED
HOUSEHOLD_ID

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 24

Strategic Partner & Systems Integrator

Intelligent Business Intelligencesm



The MIT 2008 Information Quality Industry Symposium

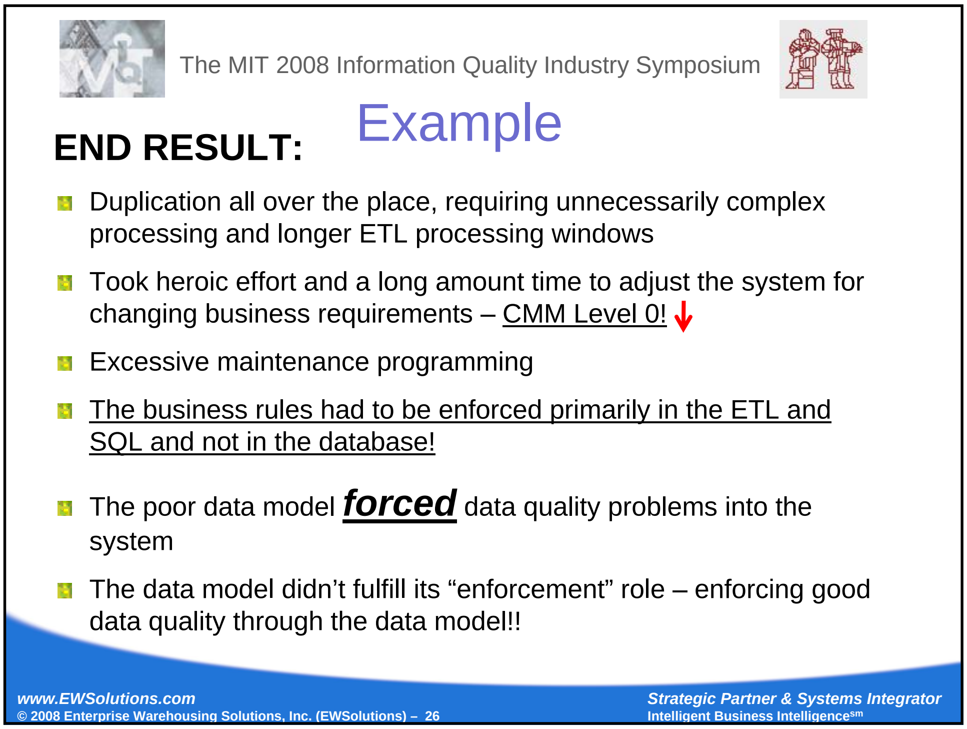
Example

FACTORS:

- Developers assumed they understood the business – they interviewed the customer
- A CDM was not created due many factors such as lack of data modeling expertise and tight deadlines.
- Was <u>incredibly</u> difficult to make changes to the model
- This "proof of concept" required very extensive modification in order for the business to have some confidence in it . It was eventually outsourced!

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 25

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

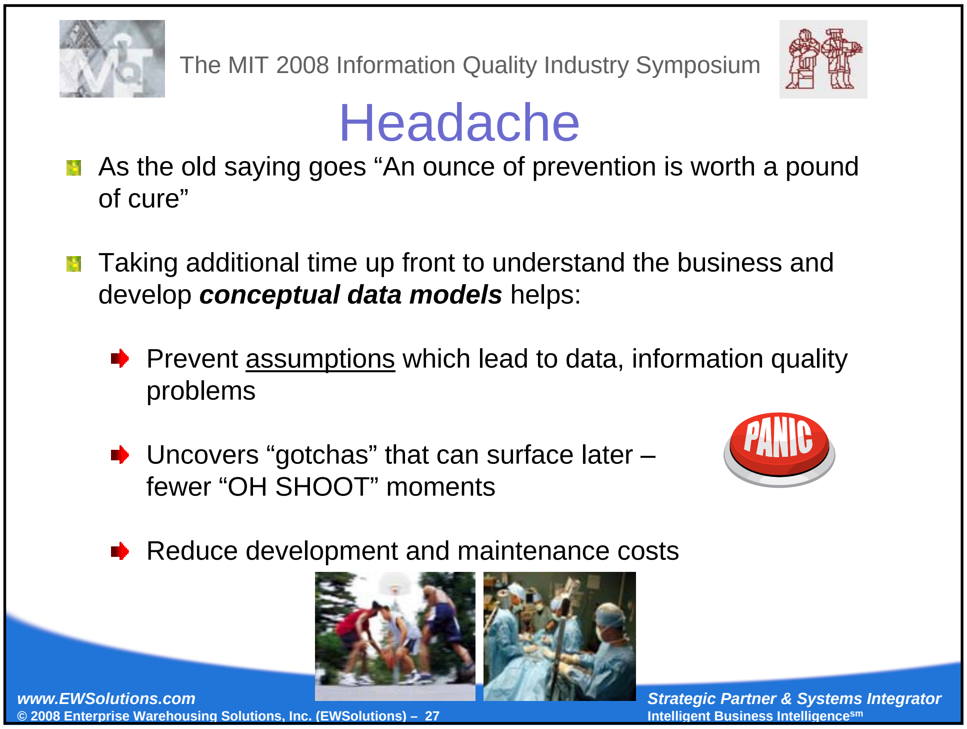
Example

END RESULT:

- Duplication all over the place, requiring unnecessarily complex processing and longer ETL processing windows
- Took heroic effort and a long amount time to adjust the system for changing business requirements – <u>CMM Level 0!</u> ↓
- Excessive maintenance programming
- <u>The business rules had to be enforced primarily in the ETL and SQL and not in the database!</u>
- The poor data model ***<u>forced</u>*** data quality problems into the system
- The data model didn't fulfill its "enforcement" role – enforcing good data quality through the data model!!

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 26

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



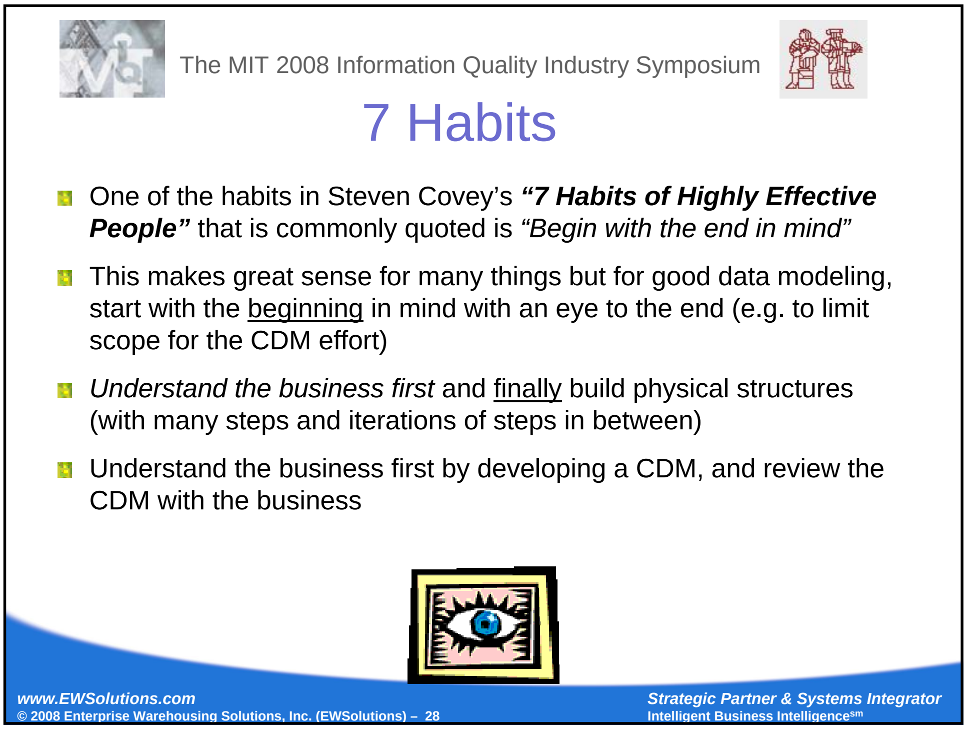
The MIT 2008 Information Quality Industry Symposium

Headache

- As the old saying goes “An ounce of prevention is worth a pound of cure”
- Taking additional time up front to understand the business and develop ***conceptual data models*** helps:
  - Prevent <u>assumptions</u> which lead to data, information quality problems
  - Uncovers “gotchas” that can surface later – fewer “OH SHOOT” moments
  - Reduce development and maintenance costs

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 27

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



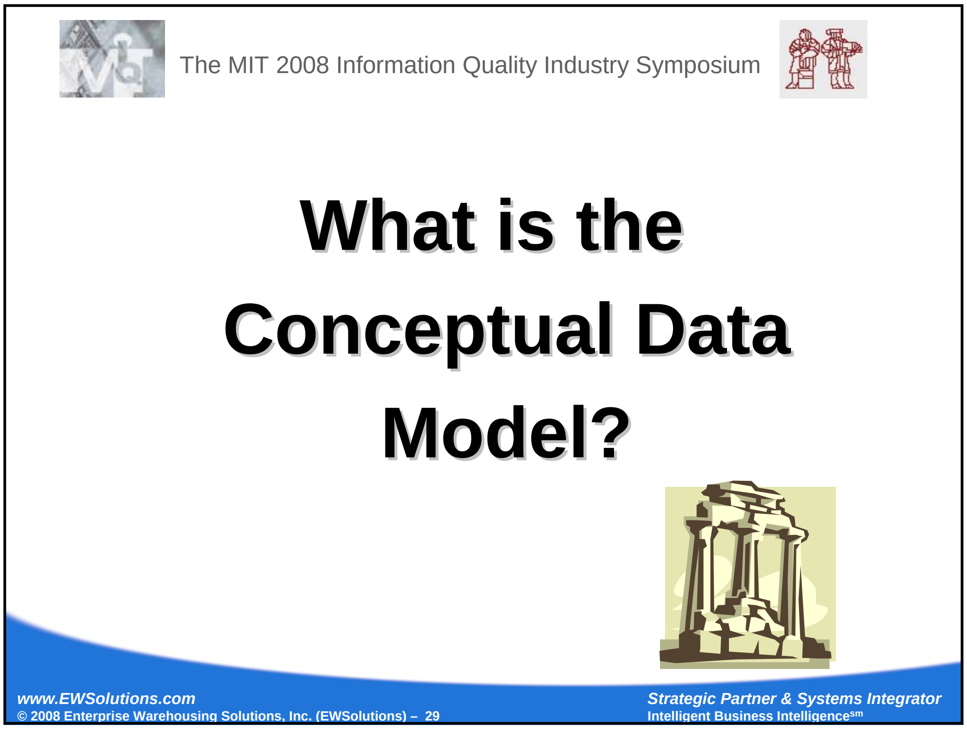
The MIT 2008 Information Quality Industry Symposium

7 Habits

- One of the habits in Steven Covey’s ***“7 Habits of Highly Effective People”*** that is commonly quoted is *“Begin with the end in mind”*
- This makes great sense for many things but for good data modeling, start with the <u>beginning</u> in mind with an eye to the end (e.g. to limit scope for the CDM effort)
- *Understand the business first* and <u>finally</u> build physical structures (with many steps and iterations of steps in between)
- Understand the business first by developing a CDM, and review the CDM with the business

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 28

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}

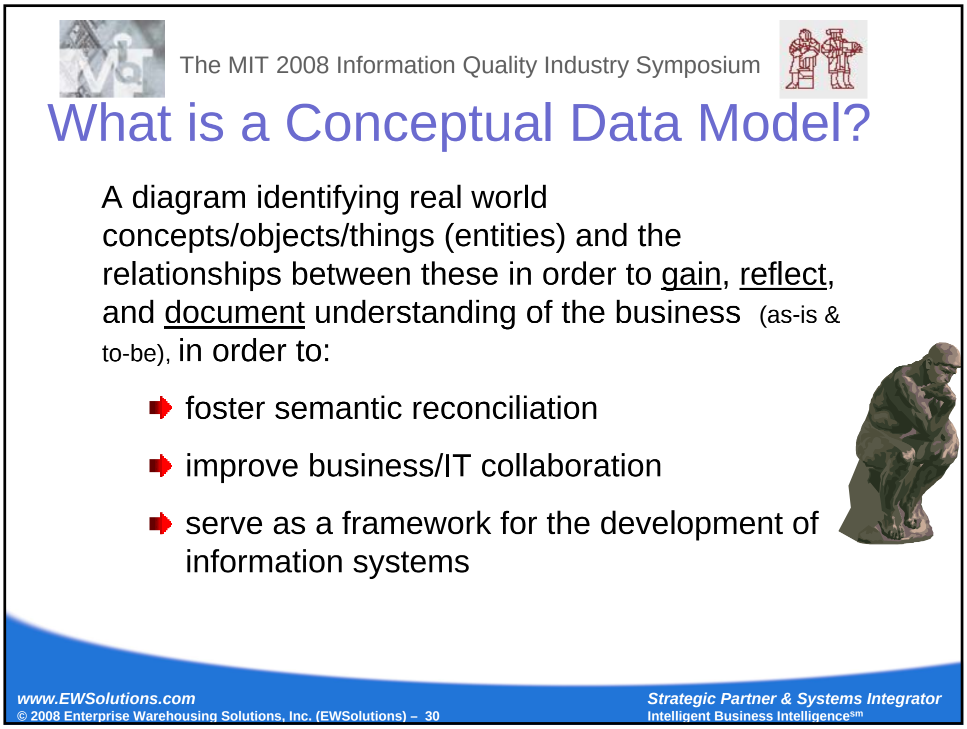


The MIT 2008 Information Quality Industry Symposium

What is the Conceptual Data Model?

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 29

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

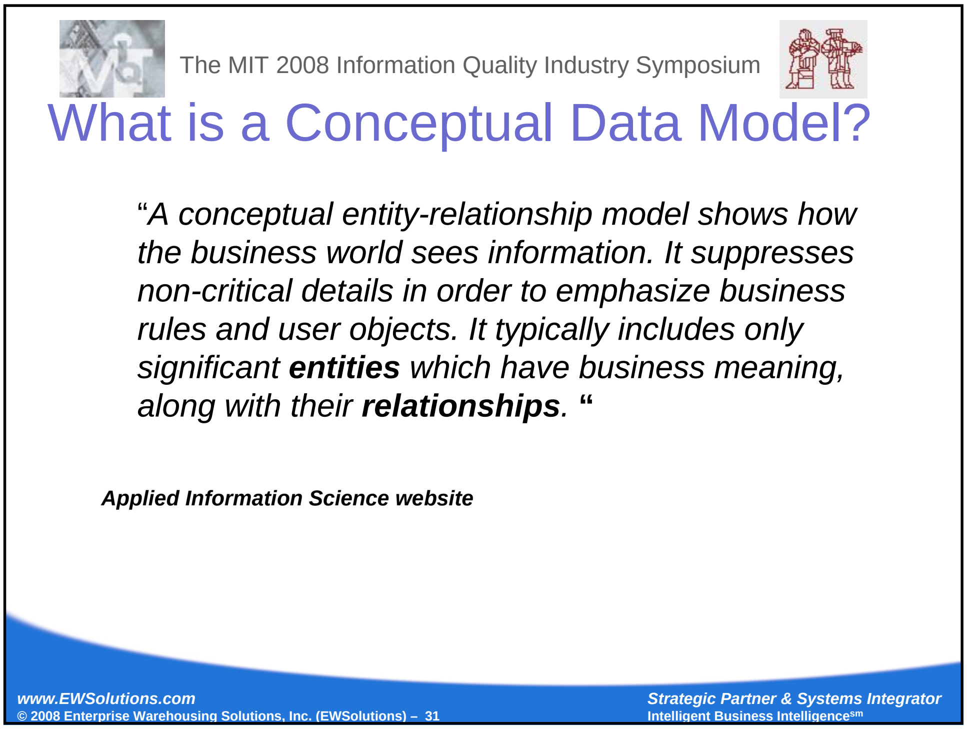
What is a Conceptual Data Model?

A diagram identifying real world concepts/objects/things (entities) and the relationships between these in order to gain, reflect, and document understanding of the business (as-is & to-be), in order to:

- foster semantic reconciliation
- improve business/IT collaboration
- serve as a framework for the development of information systems

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 30

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

What is a Conceptual Data Model?

*"A conceptual entity-relationship model shows how the business world sees information. It suppresses non-critical details in order to emphasize business rules and user objects. It typically includes only significant **entities** which have business meaning, along with their **relationships**.* **"**

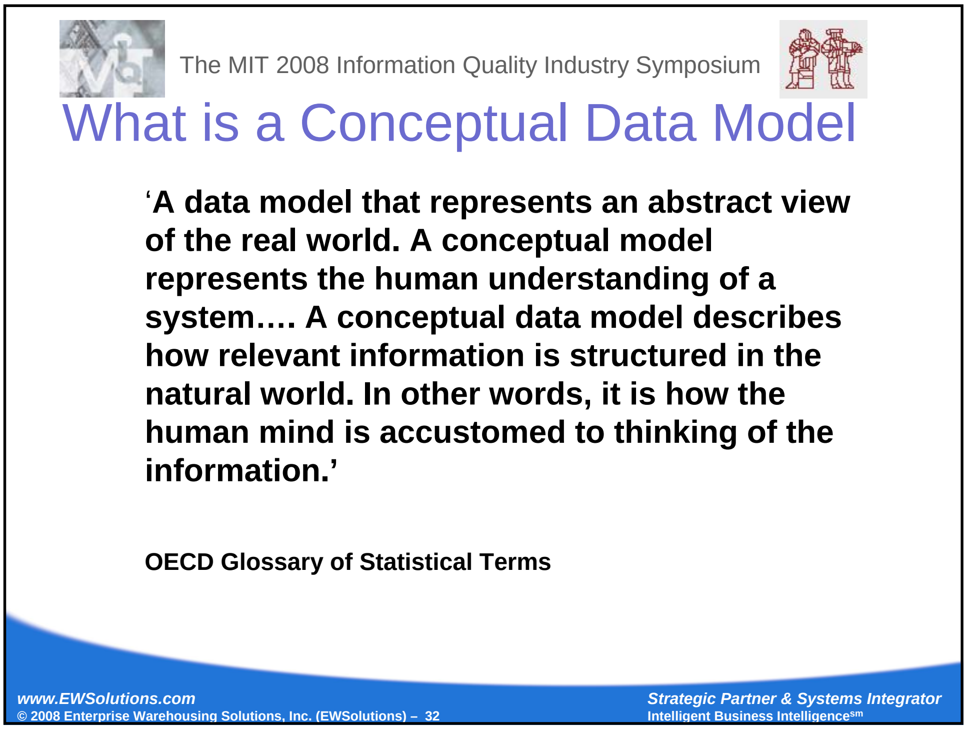
Applied Information Science website

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 31

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^sm



The MIT 2008 Information Quality Industry Symposium

What is a Conceptual Data Model

'A data model that represents an abstract view of the real world. A conceptual model represents the human understanding of a system…. A conceptual data model describes how relevant information is structured in the natural world. In other words, it is how the human mind is accustomed to thinking of the information.'

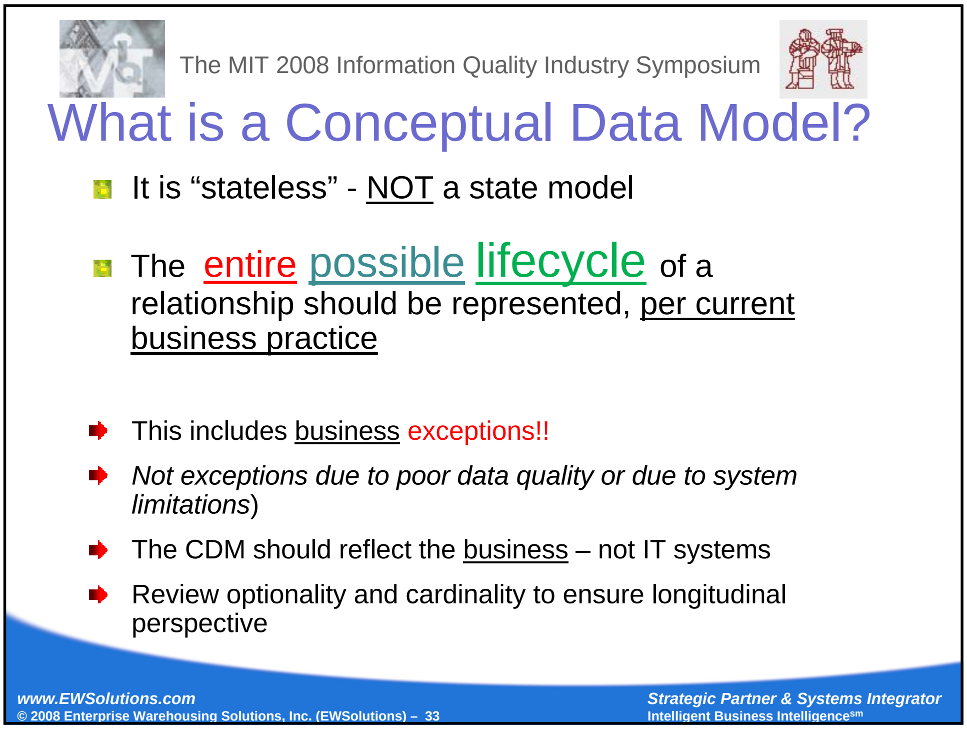
OECD Glossary of Statistical Terms

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 32

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^sm



The MIT 2008 Information Quality Industry Symposium

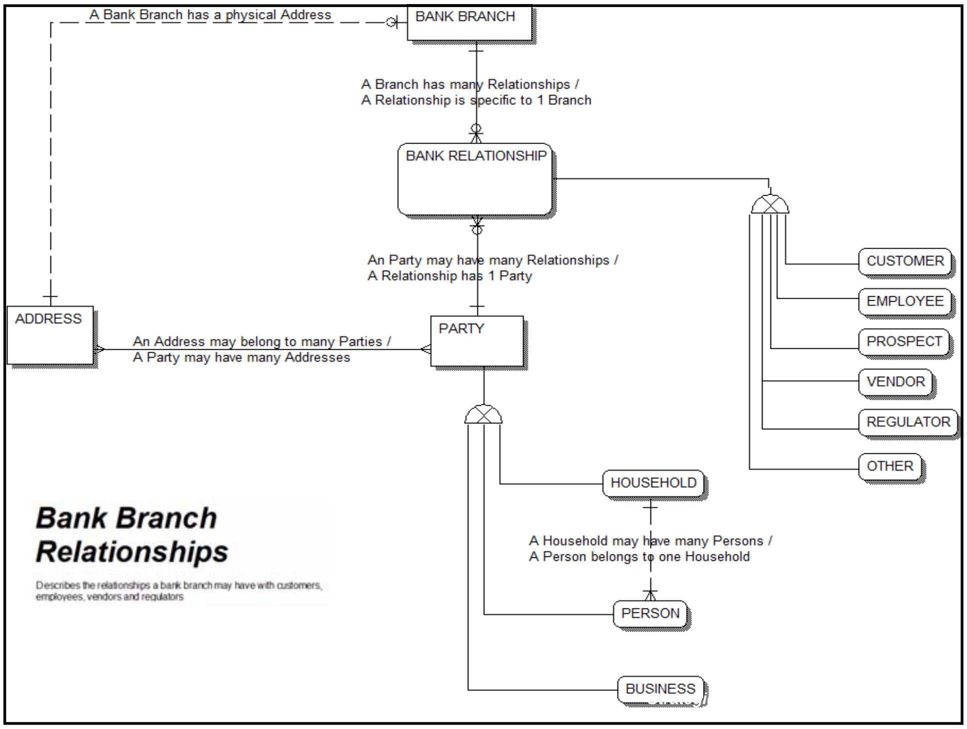
What is a Conceptual Data Model?

- It is "stateless" - NOT a state model
- The entire possible lifecycle of a relationship should be represented, per current business practice

- This includes business exceptions!!
- *Not exceptions due to poor data quality or due to system limitations*)
- The CDM should reflect the business – not IT systems
- Review optionality and cardinality to ensure longitudinal perspective

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 33

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



A Bank Branch has a physical Address

BANK BRANCH

A Branch has many Relationships /
A Relationship is specific to 1 Branch

BANK RELATIONSHIP

An Party may have many Relationships /
A Relationship has 1 Party

ADDRESS

An Address may belong to many Parties /
A Party may have many Addresses

PARTY

CUSTOMER

EMPLOYEE

PROSPECT

VENDOR

REGULATOR

OTHER

HOUSEHOLD

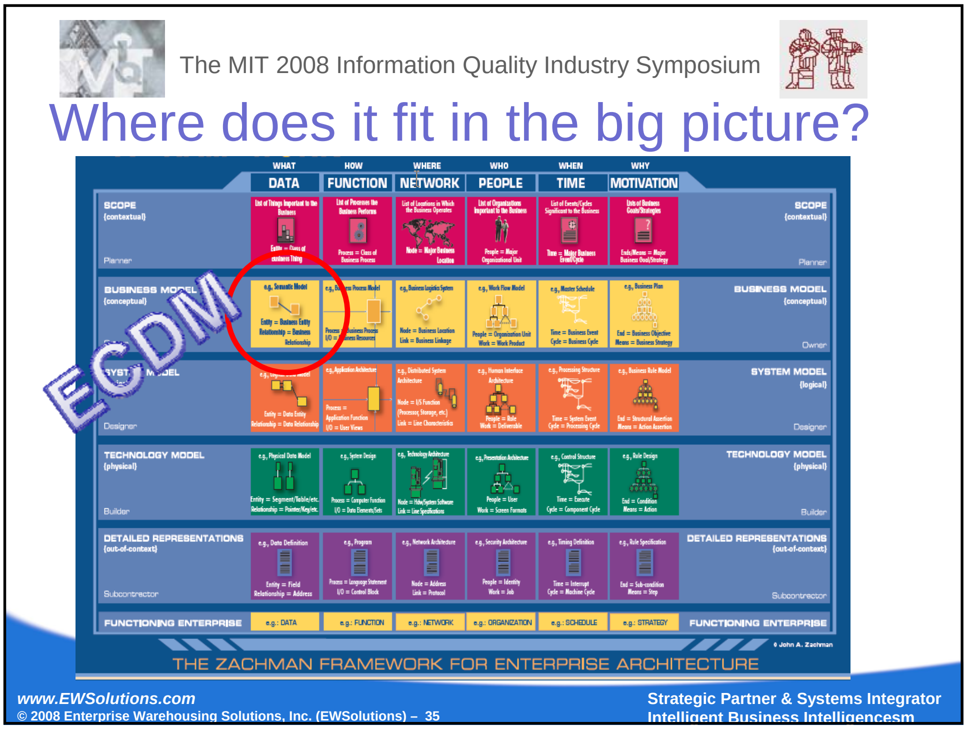
A Household may have many Persons /
A Person belongs to one Household

PERSON

BUSINESS

Bank Branch Relationships

Describes the relationships a bank branch may have with customers, employees, vendors and regulators



The MIT 2008 Information Quality Industry Symposium

Where does it fit in the big picture?

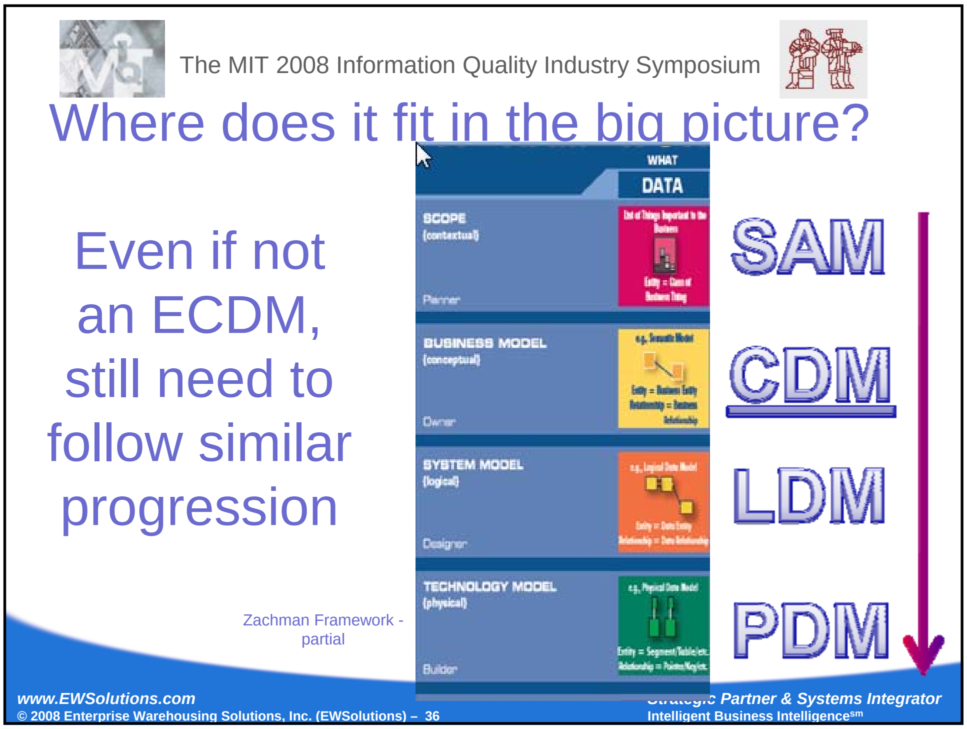
	WHAT DATA	HOW FUNCTION	WHERE NETWORK	WHO PEOPLE	WHEN TIME	WHY MOTIVATION	
SCOPE {contextual} Planner	List of Things Important to the Business Entity = Class of Business Thing	List of Processes the Business Performs Process = Class of Business Process	List of Locations in Which the Business Operates Node = Major Business Location	List of Organizations Important to the Business People = Major Organizational Unit	List of Events/Cycles Significant to the Business Time = Major Business Event/Cycle	Lists of Business Goals/Strategies Ends/Means = Major Business Goal/Strategy	SCOPE {contextual} Planner
BUSINESS MODEL {conceptual} Owner	e.g., Semantic Model Entity = Business Entity Relationship = Business Relationship	e.g., Business Process Model Process = Business Process I/O = Business Resources	e.g., Business Logistics System Node = Business Location Link = Business Linkage	e.g., Work Flow Model People = Organization Unit Work = Work Product	e.g., Master Schedule Time = Business Event Cycle = Business Cycle	e.g., Business Plan End = Business Objective Means = Business Strategy	BUSINESS MODEL {conceptual} Owner
SYSTEM MODEL {logical} Designer	e.g., Logical Data Model Entity = Data Entity Relationship = Data Relationship	e.g., Application Architecture Process = Application Function I/O = User Views	e.g., Distributed System Architecture Node = I/S Function (Processor, Storage, etc.) Link = Line Characteristics	e.g., Human Interface Architecture People = Role Work = Deliverable	e.g., Processing Structure Time = System Event Cycle = Processing Cycle	e.g., Business Rule Model End = Structural Assertion Means = Action Assertion	SYSTEM MODEL {logical} Designer
TECHNOLOGY MODEL {physical} Builder	e.g., Physical Data Model Entity = Segment/Table/etc. Relationship = Pointer/Key/etc.	e.g., System Design Process = Computer Function I/O = Data Elements/Sets	e.g., Technology Architecture Node = Hdw/System Software Link = Line Specifications	e.g., Presentation Architecture People = User Work = Screen Formats	e.g., Control Structure Time = Execute Cycle = Component Cycle	e.g., Rule Design End = Condition Means = Action	TECHNOLOGY MODEL {physical} Builder
DETAILED REPRESENTATIONS {out-of-context} Subcontractor	e.g., Data Definition Entity = Field Relationship = Address	e.g., Program Process = Language Statement I/O = Control Block	e.g., Network Architecture Node = Address Link = Protocol	e.g., Security Architecture People = Identity Work = Job	e.g., Timing Definition Time = Interrupt Cycle = Machine Cycle	e.g., Rule Specification End = Sub-condition Means = Step	DETAILED REPRESENTATIONS {out-of-context} Subcontractor
FUNCTIONING ENTERPRISE	e.g.: DATA	e.g.: FUNCTION	e.g.: NETWORK	e.g.: ORGANIZATION	e.g.: SCHEDULE	e.g.: STRATEGY	FUNCTIONING ENTERPRISE

© John A. Zachman

THE ZACHMAN FRAMEWORK FOR ENTERPRISE ARCHITECTURE

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 35

Strategic Partner & Systems Integrator
Intelligent Business Intelligencesm



The MIT 2008 Information Quality Industry Symposium

Where does it fit in the big picture?

Even if not an ECDM, still need to follow similar progression

	WHAT DATA
SCOPE (contextual) Planner	List of Things Important to the Business Entity = Class of Business Thing

BUSINESS MODEL (conceptual)
Owner
	e.g., Semantic Model
	Entity = Business Entity
	Relationship = Business Relationship

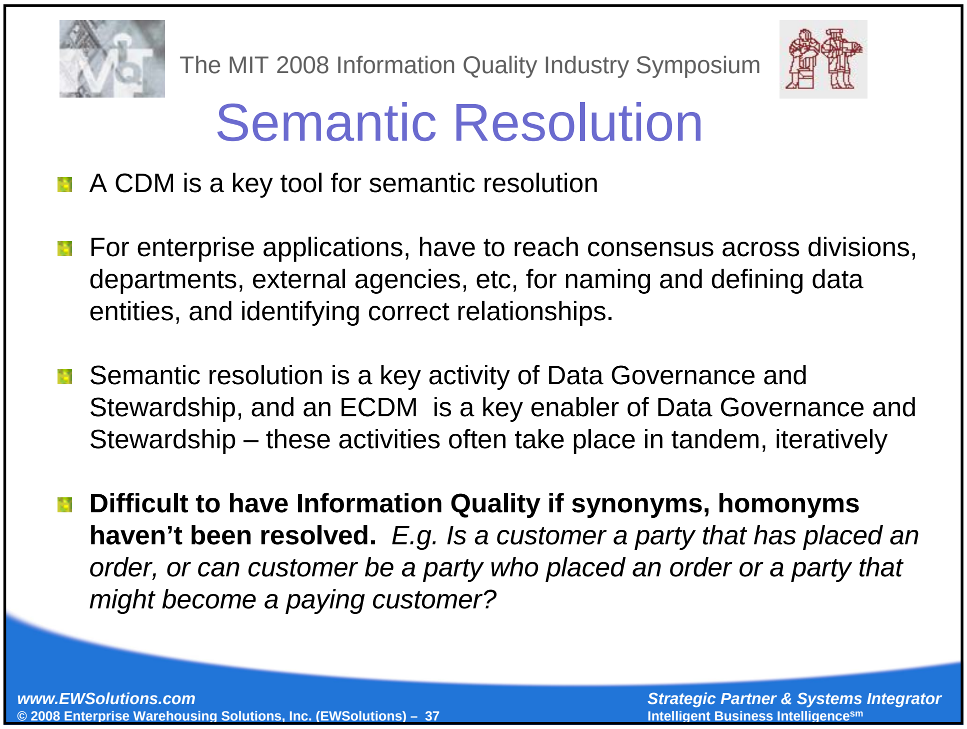
SYSTEM MODEL (logical)
Designer
	e.g., Logical Data Model
	Entity = Data Entity
	Relationship = Data Relationship

TECHNOLOGY MODEL (physical)
Builder
	e.g., Physical Data Model
	Entity = Segment/Table/etc.
	Relationship = Pointer/Key/etc.

Zachman Framework - partial

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 36

Strategic Partner & Systems Integrator
Intelligent Business Intelligencesm



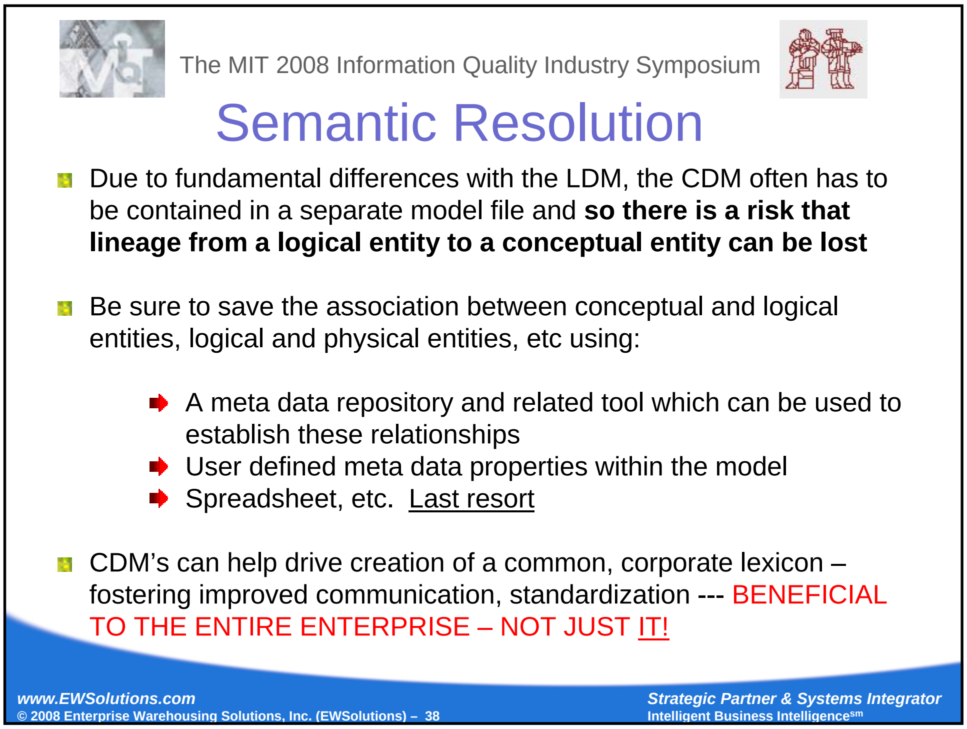
The MIT 2008 Information Quality Industry Symposium

Semantic Resolution

- A CDM is a key tool for semantic resolution
- For enterprise applications, have to reach consensus across divisions, departments, external agencies, etc, for naming and defining data entities, and identifying correct relationships.
- Semantic resolution is a key activity of Data Governance and Stewardship, and an ECDM is a key enabler of Data Governance and Stewardship – these activities often take place in tandem, iteratively
- **Difficult to have Information Quality if synonyms, homonyms haven't been resolved.** *E.g. Is a customer a party that has placed an order, or can customer be a party who placed an order or a party that might become a paying customer?*

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 37

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



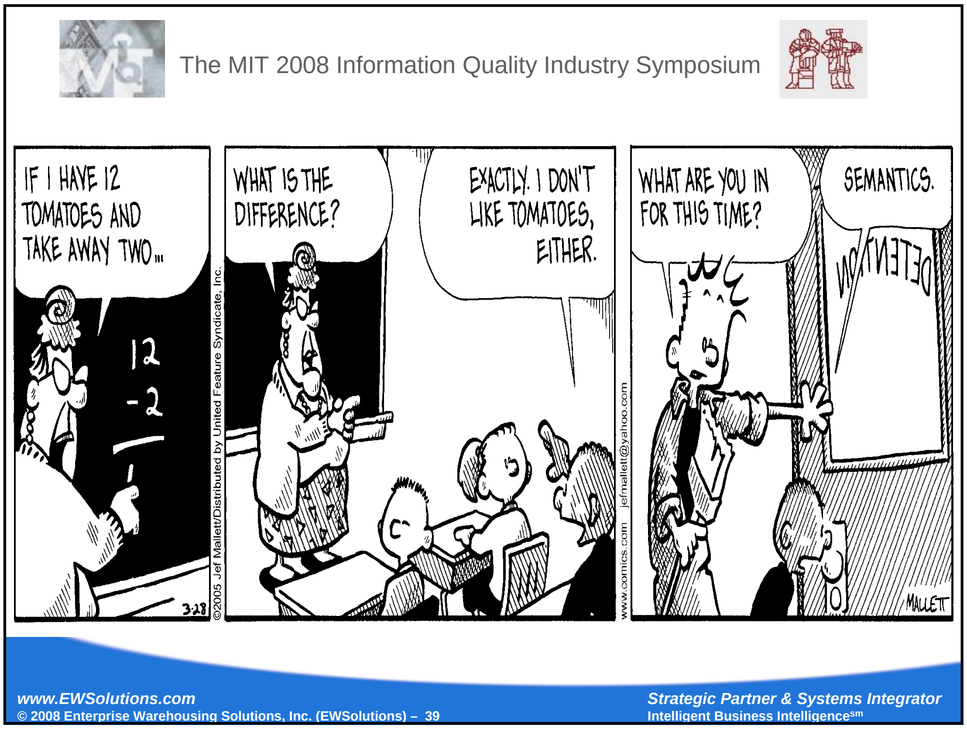
The MIT 2008 Information Quality Industry Symposium

Semantic Resolution

- Due to fundamental differences with the LDM, the CDM often has to be contained in a separate model file and **so there is a risk that lineage from a logical entity to a conceptual entity can be lost**
- Be sure to save the association between conceptual and logical entities, logical and physical entities, etc using:
  - A meta data repository and related tool which can be used to establish these relationships
  - User defined meta data properties within the model
  - Spreadsheet, etc. Last resort
- CDM's can help drive creation of a common, corporate lexicon – fostering improved communication, standardization --- BENEFICIAL TO THE ENTIRE ENTERPRISE – NOT JUST IT!

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 38

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

IF I HAVE 12 TOMATOES AND TAKE AWAY TWO...
12
-2
WHAT IS THE DIFFERENCE?
EXACTLY. I DON'T LIKE TOMATOES, EITHER.
WHAT ARE YOU IN FOR THIS TIME?
SEMANTICS.
DETENTION
3-28
MALLETT

©2005 Jef Mallett/Distributed by United Feature Syndicate, Inc.

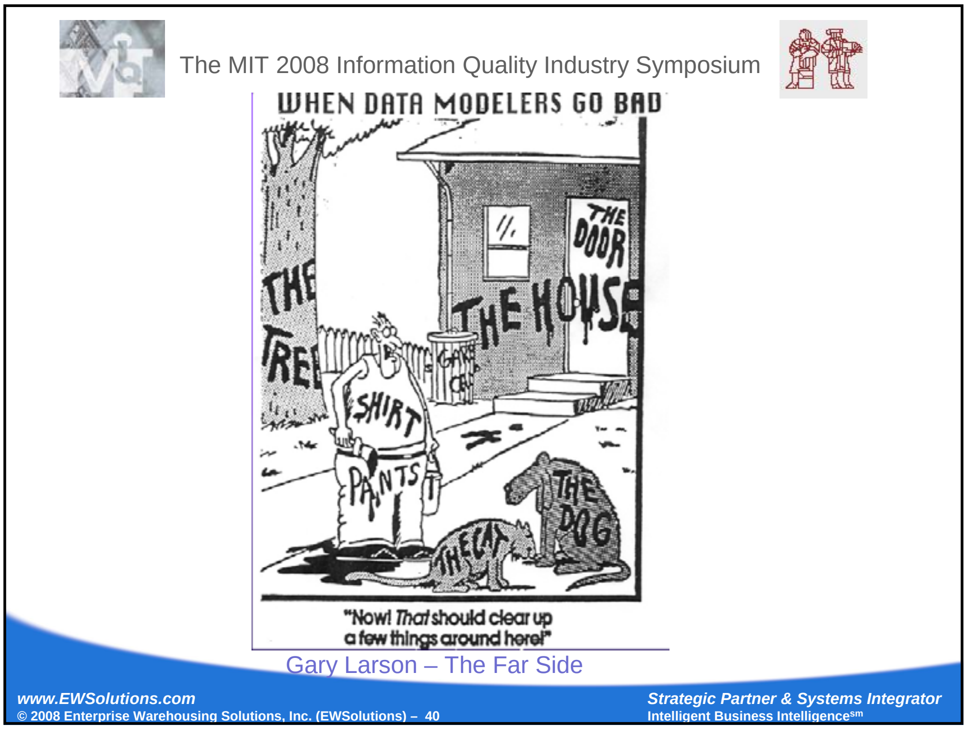
www.comics.com jefmallett@yahoo.com

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 39

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

"Now! That should clear up a few things around here!"

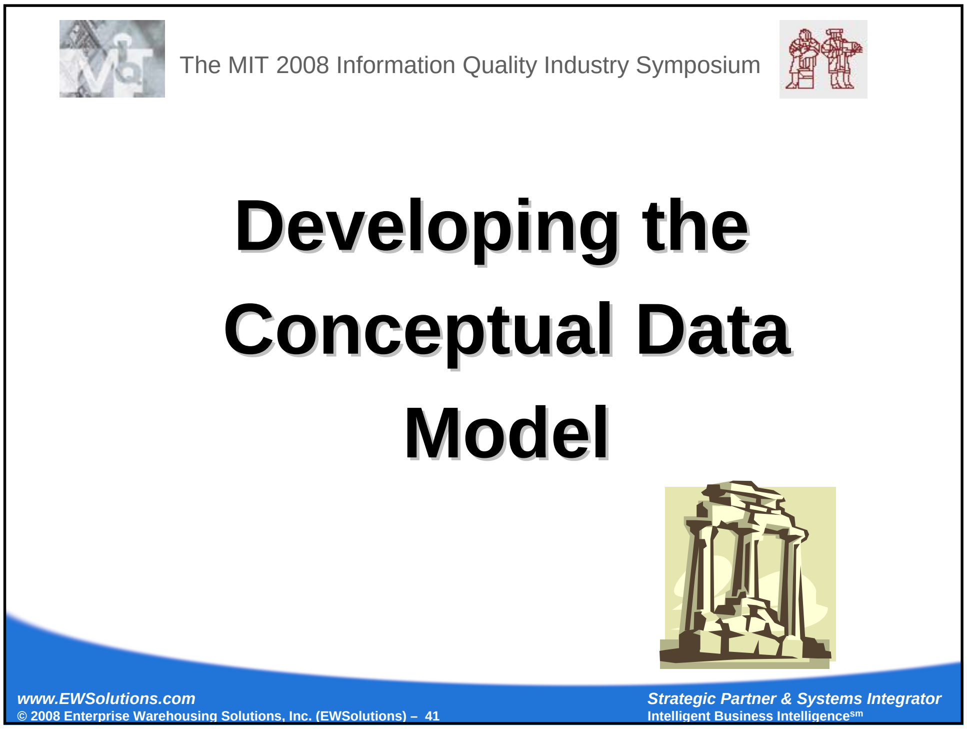
Gary Larson – The Far Side

www.EWSolutions.com

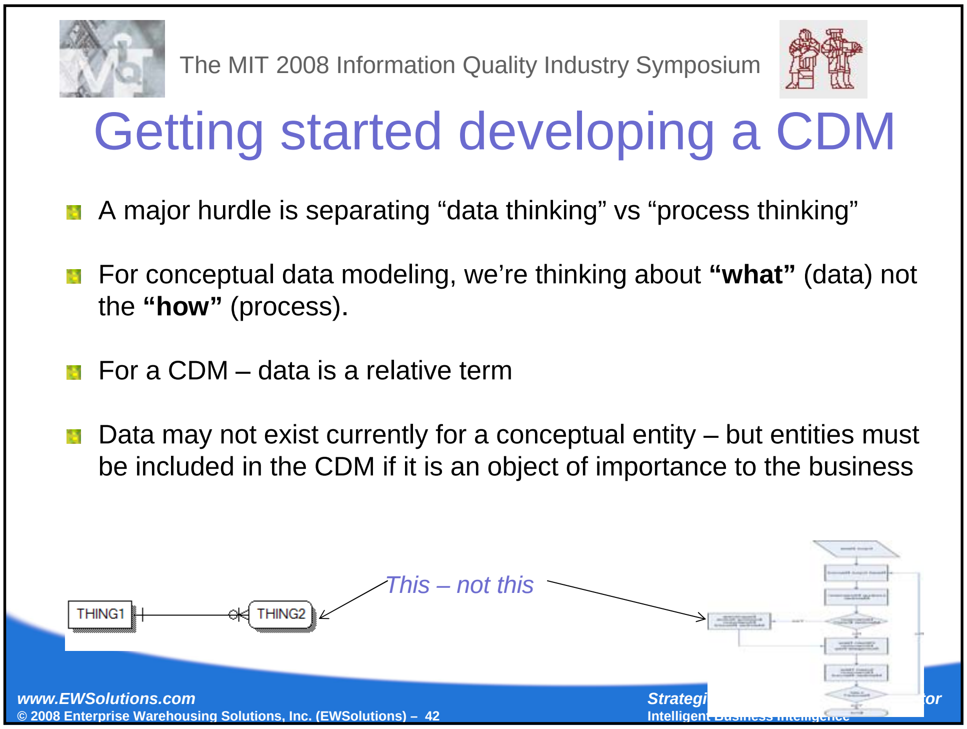
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 40

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



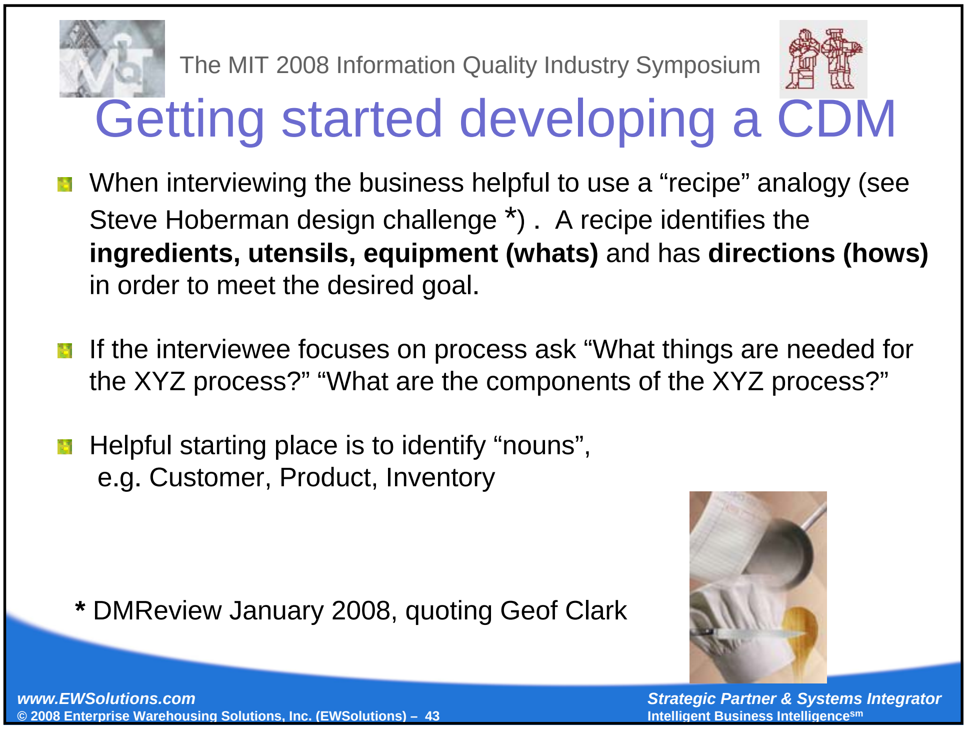
Developing the Conceptual Data Model



Getting started developing a CDM

- A major hurdle is separating “data thinking” vs “process thinking”
- For conceptual data modeling, we’re thinking about **“what”** (data) not the **“how”** (process).
- For a CDM – data is a relative term
- Data may not exist currently for a conceptual entity – but entities must be included in the CDM if it is an object of importance to the business

This – not this



The MIT 2008 Information Quality Industry Symposium

Getting started developing a CDM

- When interviewing the business helpful to use a “recipe” analogy (see Steve Hoberman design challenge *) . A recipe identifies the **ingredients, utensils, equipment (whats)** and has **directions (hows)** in order to meet the desired goal.

- If the interviewee focuses on process ask “What things are needed for the XYZ process?” “What are the components of the XYZ process?”

- Helpful starting place is to identify “nouns”, e.g. Customer, Product, Inventory

***** DMReview January 2008, quoting Geof Clark

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 43

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

The MIT 2008 Information Quality Industry Symposium

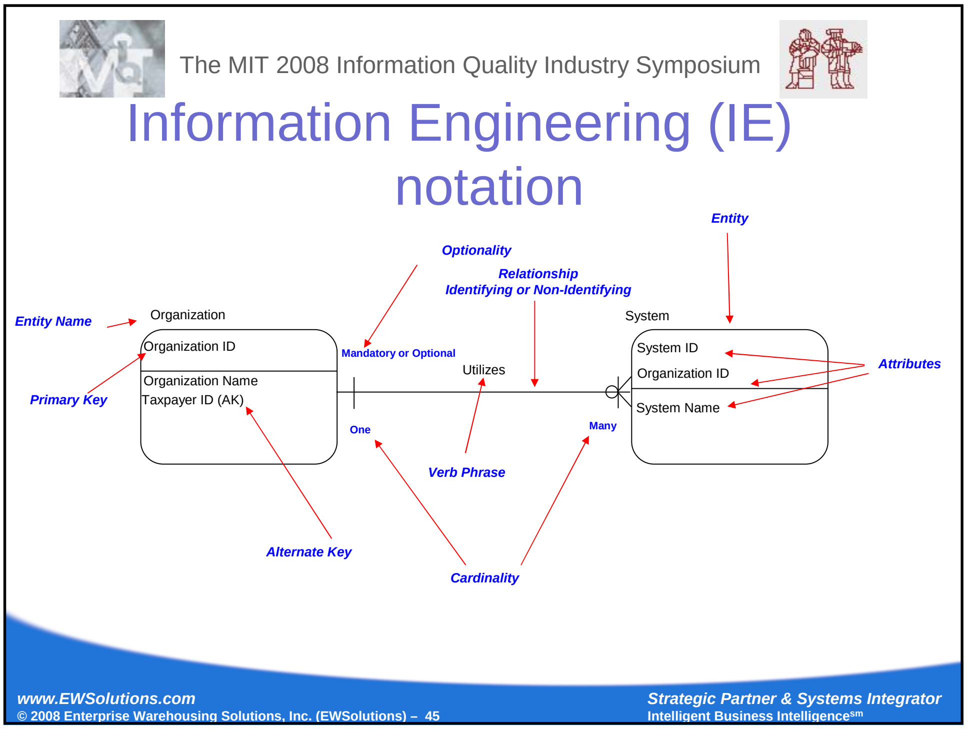
When is a CDM finished?

“Perfection does not come into being, when nothing more can be added, but when nothing can be taken away”

Antoine de Saint-Exupéry

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 44

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

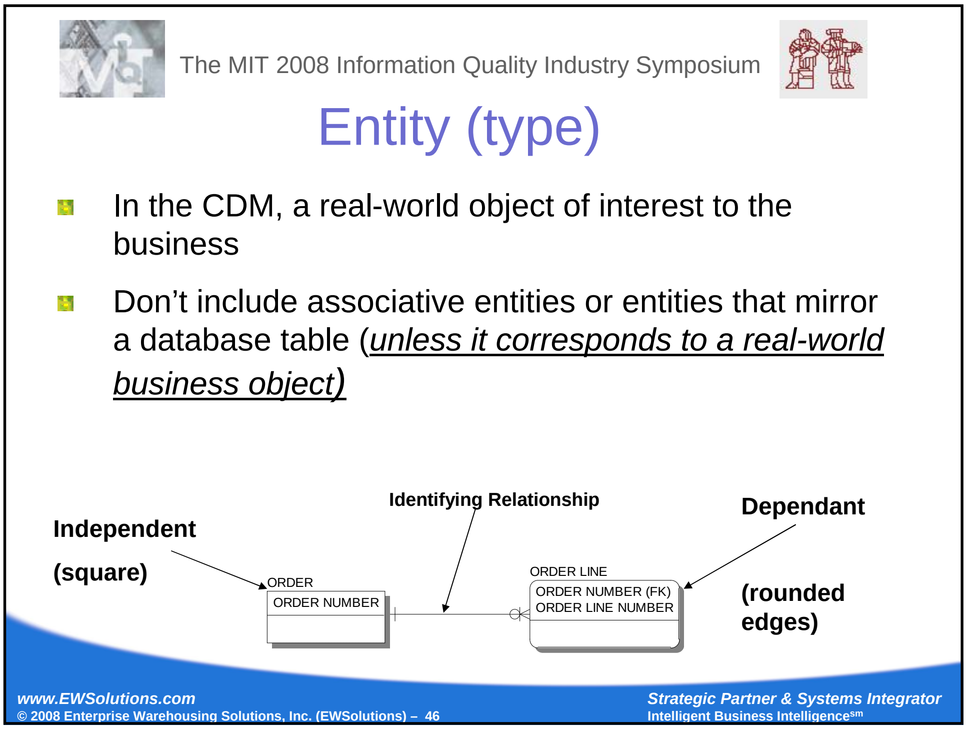


The MIT 2008 Information Quality Industry Symposium

Information Engineering (IE) notation

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 45

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

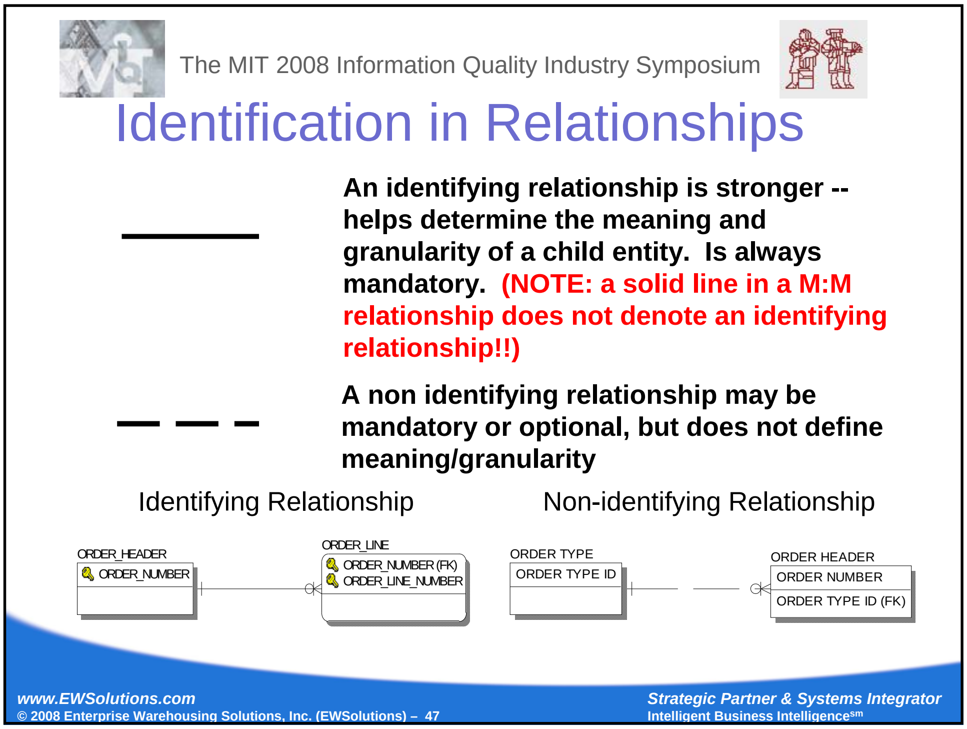
Entity (type)

- In the CDM, a real-world object of interest to the business
- Don’t include associative entities or entities that mirror a database table (*__unless it corresponds to a real-world business object__*)

Identifying Relationship
Independent
(square)
Dependant
ORDER
ORDER NUMBER
ORDER LINE
ORDER NUMBER (FK)
ORDER LINE NUMBER
(rounded
edges)

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 46

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

Identification in Relationships

An identifying relationship is stronger -- helps determine the meaning and granularity of a child entity. Is always mandatory. (NOTE: a solid line in a M:M relationship does not denote an identifying relationship!!)

A non identifying relationship may be mandatory or optional, but does not define meaning/granularity

Identifying Relationship

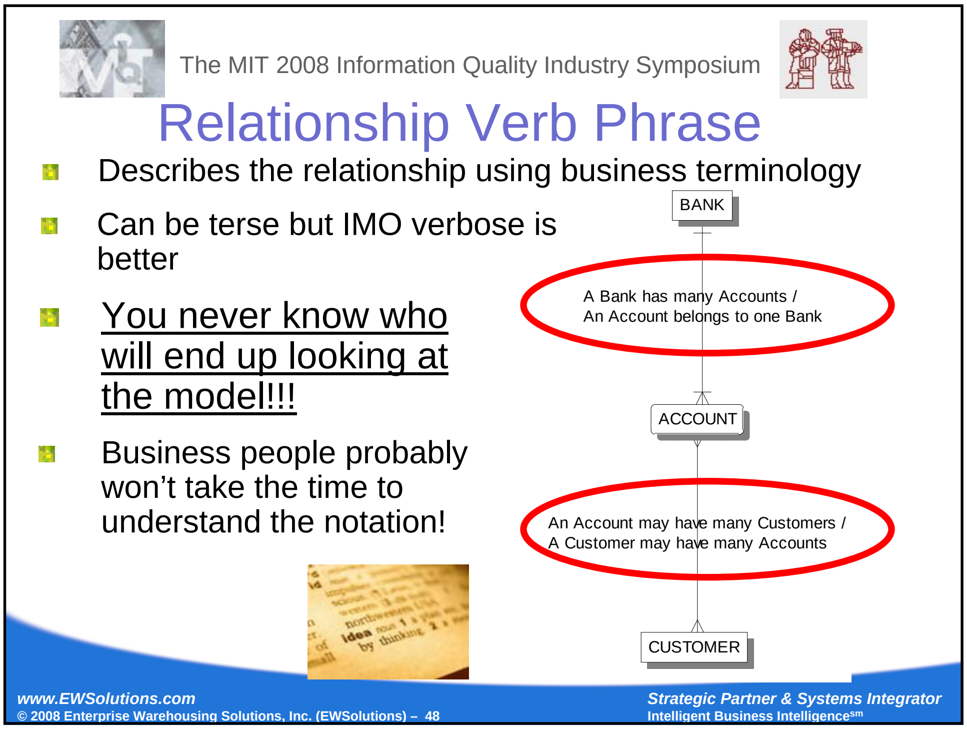
Non-identifying Relationship

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 47

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Relationship Verb Phrase

- Describes the relationship using business terminology
- Can be terse but IMO verbose is better
- <u>You never know who will end up looking at the model!!!</u>
- Business people probably won't take the time to understand the notation!

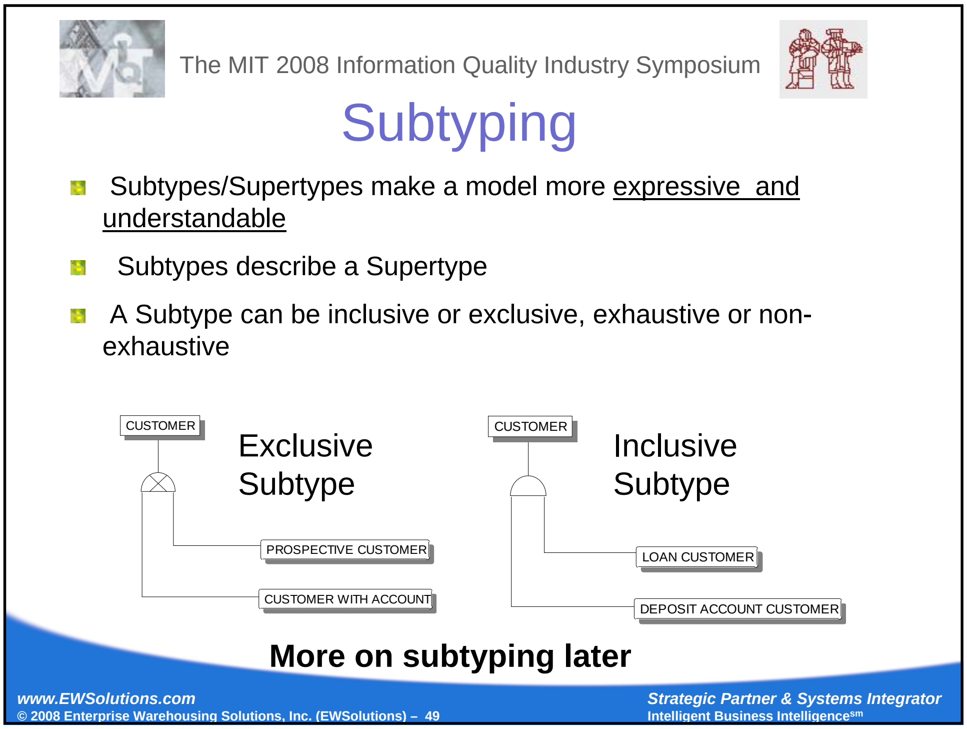
BANK
A Bank has many Accounts /
An Account belongs to one Bank
ACCOUNT
An Account may have many Customers /
A Customer may have many Accounts
CUSTOMER

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 48

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Subtyping

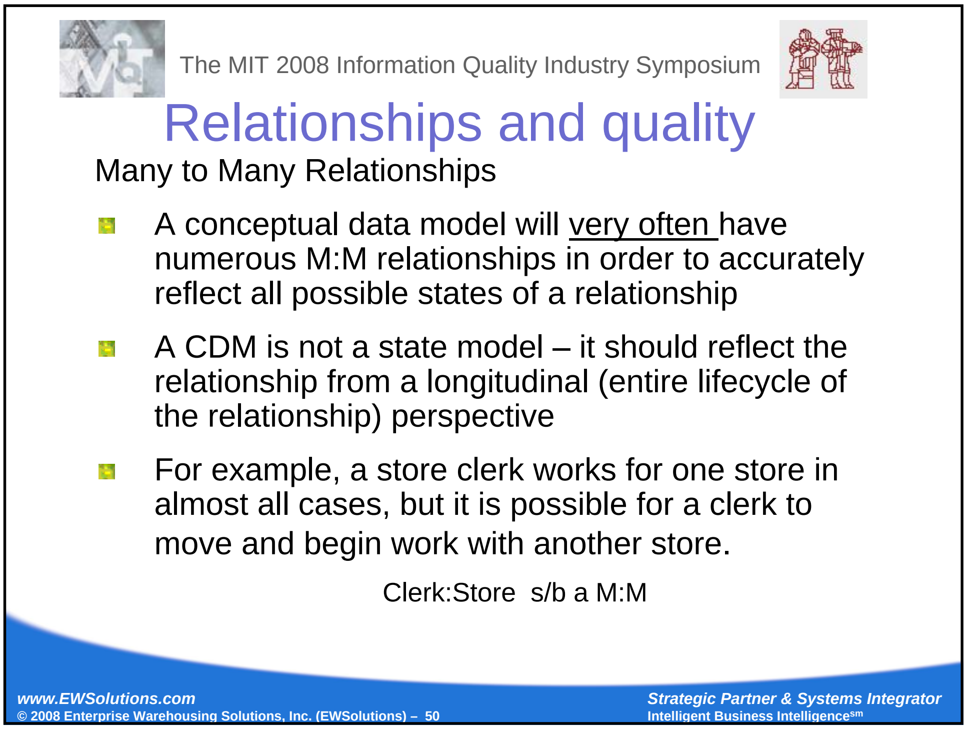
- Subtypes/Supertypes make a model more expressive and understandable
- Subtypes describe a Supertype
- A Subtype can be inclusive or exclusive, exhaustive or non-exhaustive

CUSTOMER
Exclusive Subtype
PROSPECTIVE CUSTOMER
CUSTOMER WITH ACCOUNT
CUSTOMER
Inclusive Subtype
LOAN CUSTOMER
DEPOSIT ACCOUNT CUSTOMER

More on subtyping later

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 49

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

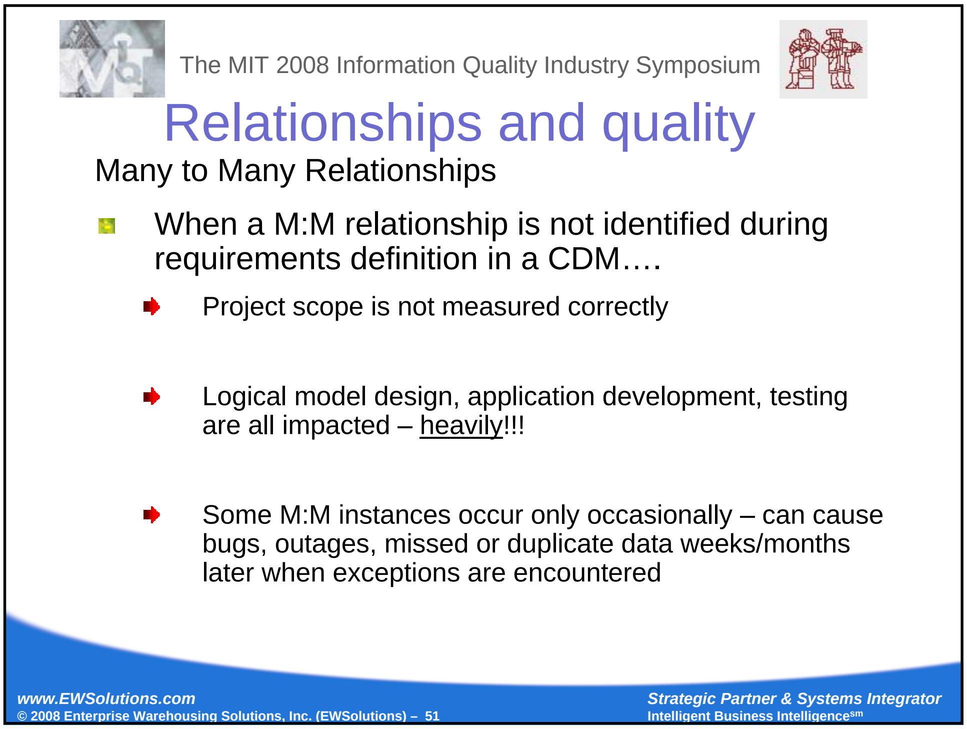
Many to Many Relationships

- A conceptual data model will very often have numerous M:M relationships in order to accurately reflect all possible states of a relationship
- A CDM is not a state model – it should reflect the relationship from a longitudinal (entire lifecycle of the relationship) perspective
- For example, a store clerk works for one store in almost all cases, but it is possible for a clerk to move and begin work with another store.

Clerk:Store s/b a M:M

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 50

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

Many to Many Relationships

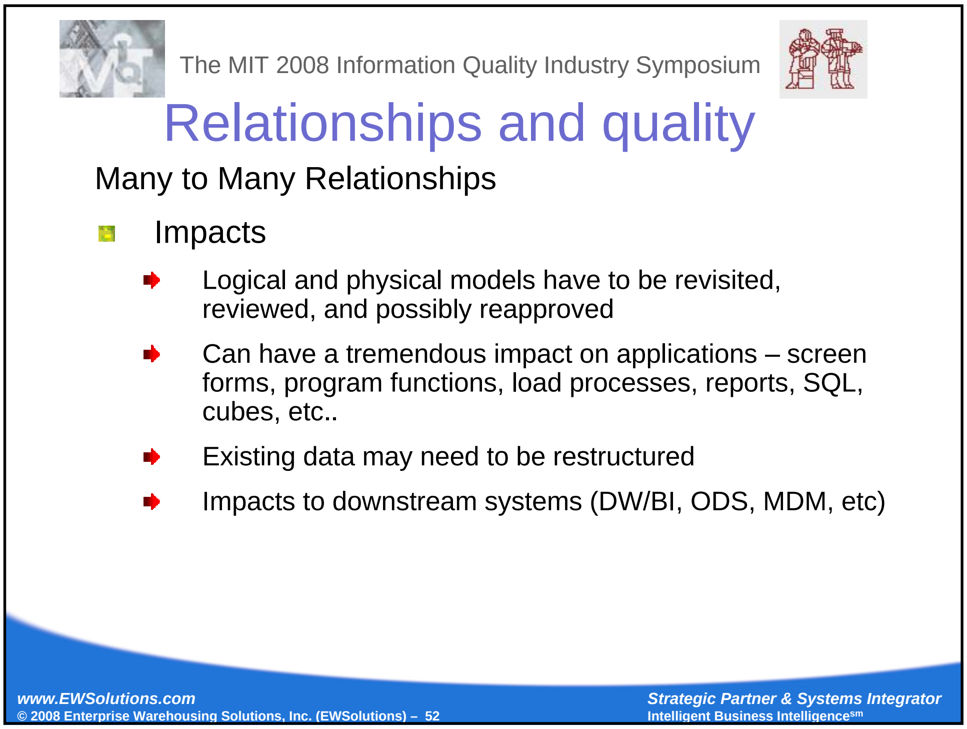
- When a M:M relationship is not identified during requirements definition in a CDM….
  - Project scope is not measured correctly
  - Logical model design, application development, testing are all impacted – heavily!!!
  - Some M:M instances occur only occasionally – can cause bugs, outages, missed or duplicate data weeks/months later when exceptions are encountered

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 51

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

Many to Many Relationships

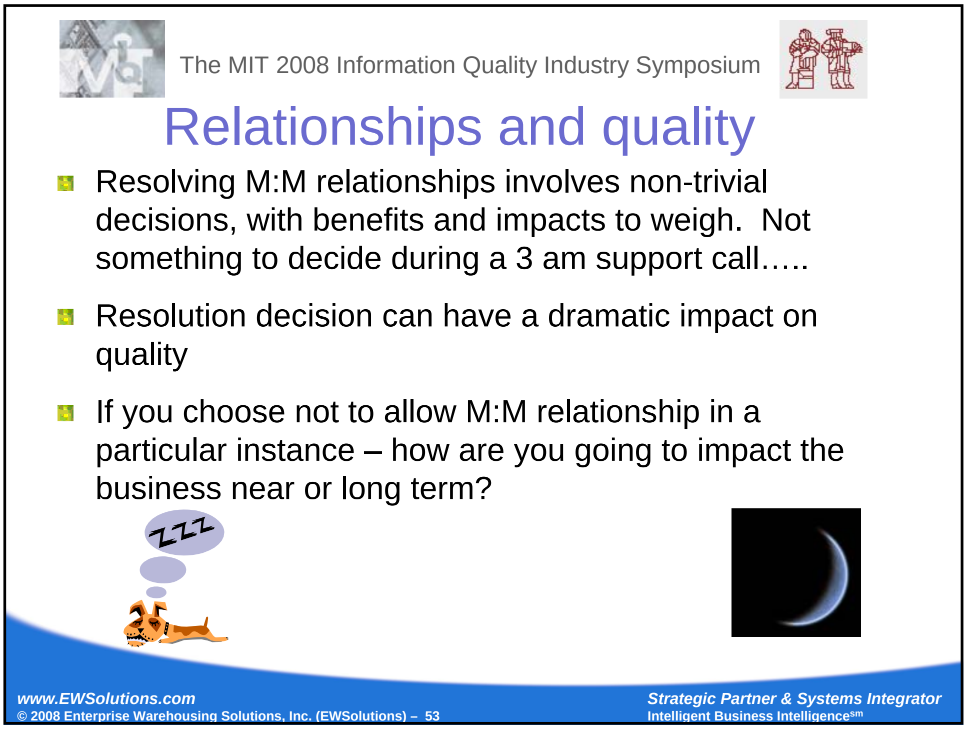
- Impacts
  - Logical and physical models have to be revisited, reviewed, and possibly reapproved
  - Can have a tremendous impact on applications – screen forms, program functions, load processes, reports, SQL, cubes, etc..
  - Existing data may need to be restructured
  - Impacts to downstream systems (DW/BI, ODS, MDM, etc)

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 52

Strategic Partner & Systems Integrator

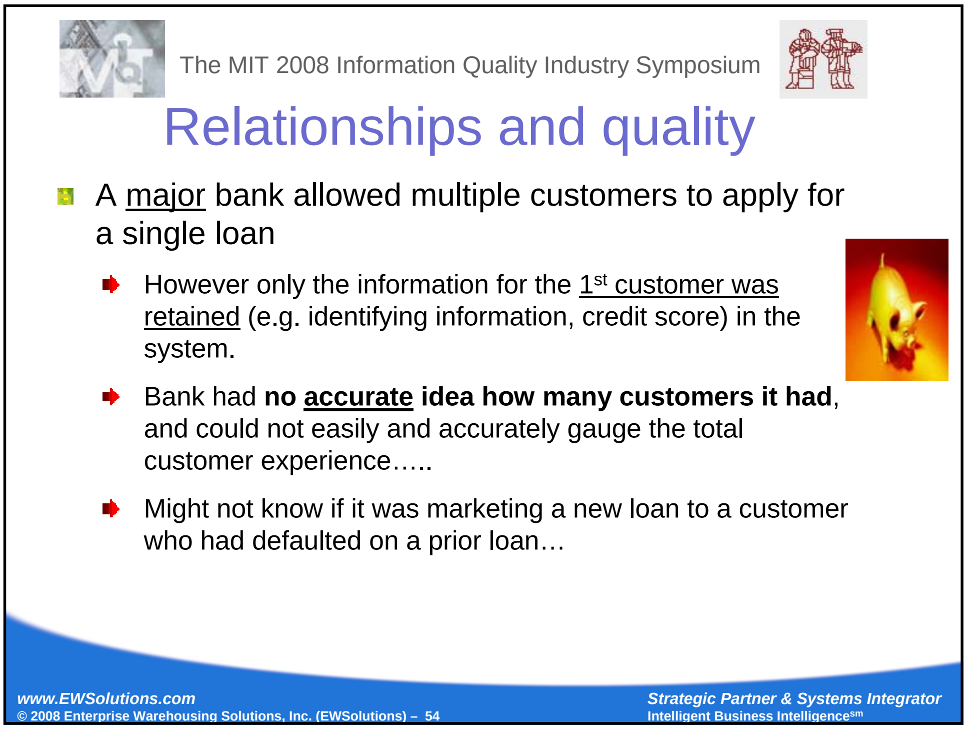
Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

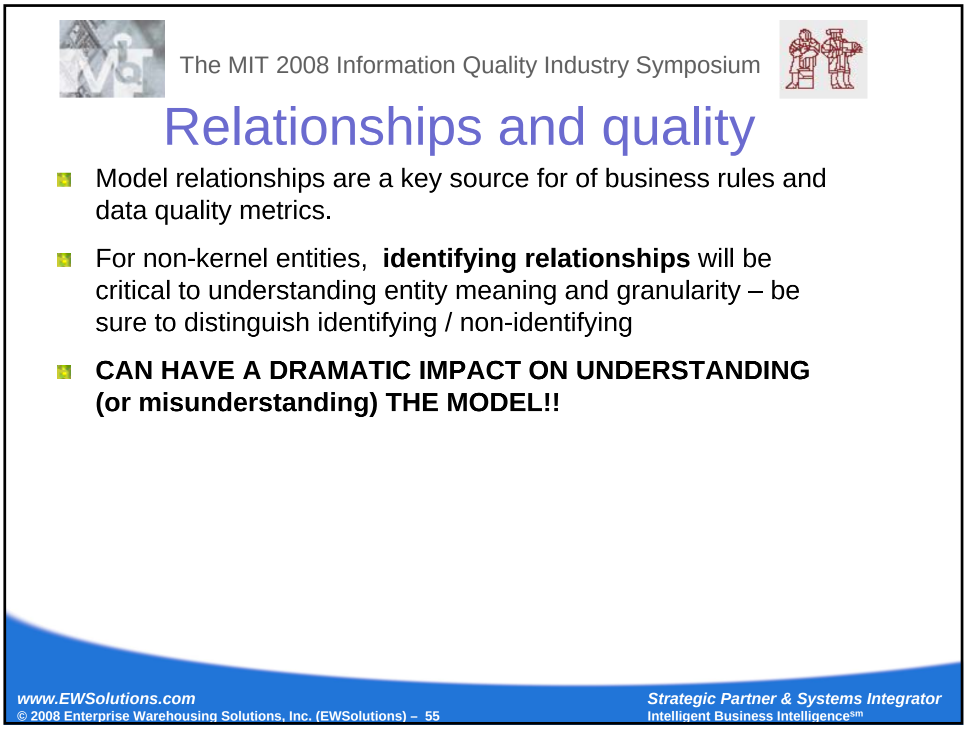
- Resolving M:M relationships involves non-trivial decisions, with benefits and impacts to weigh. Not something to decide during a 3 am support call…..
- Resolution decision can have a dramatic impact on quality
- If you choose not to allow M:M relationship in a particular instance – how are you going to impact the business near or long term?



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

- A major bank allowed multiple customers to apply for a single loan
  - However only the information for the 1st customer was retained (e.g. identifying information, credit score) in the system.
  - Bank had **no accurate idea how many customers it had**, and could not easily and accurately gauge the total customer experience…..
  - Might not know if it was marketing a new loan to a customer who had defaulted on a prior loan…



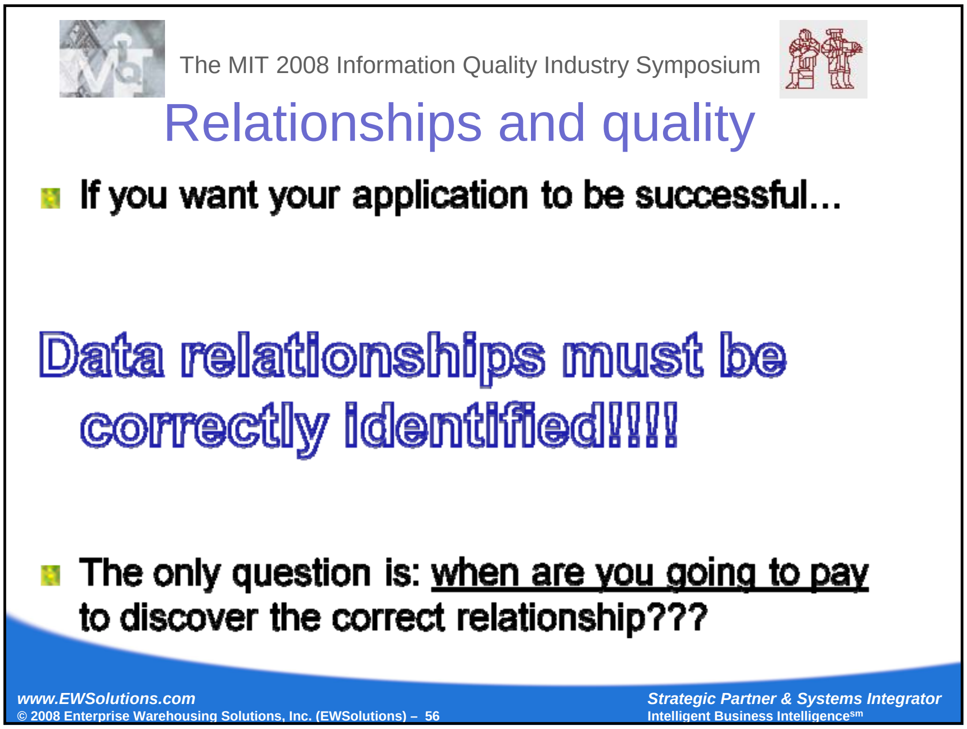
The MIT 2008 Information Quality Industry Symposium

Relationships and quality

- Model relationships are a key source for of business rules and data quality metrics.
- For non-kernel entities, **identifying relationships** will be critical to understanding entity meaning and granularity – be sure to distinguish identifying / non-identifying
- **CAN HAVE A DRAMATIC IMPACT ON UNDERSTANDING (or misunderstanding) THE MODEL!!**

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 55

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^sm



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

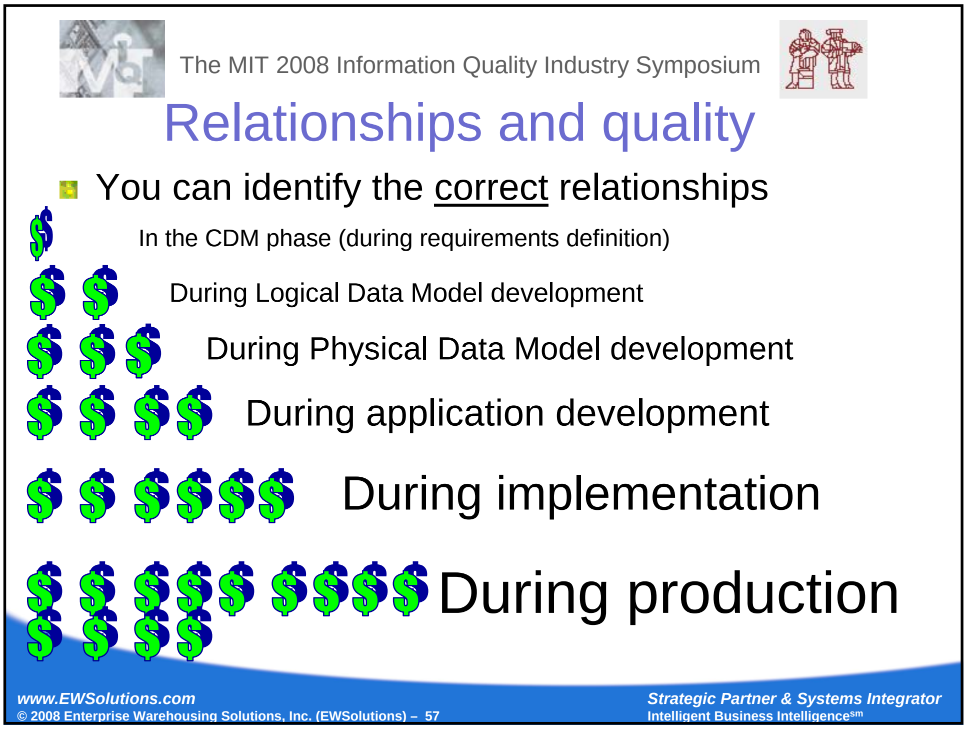
- **If you want your application to be successful…**

Data relationships must be correctly identified!!!!

- **The only question is: <u>when are you going to pay</u> to discover the correct relationship???**

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 56

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^sm



The MIT 2008 Information Quality Industry Symposium

Relationships and quality

- You can identify the correct relationships

$ In the CDM phase (during requirements definition)

$ $ During Logical Data Model development

$ $ $ During Physical Data Model development

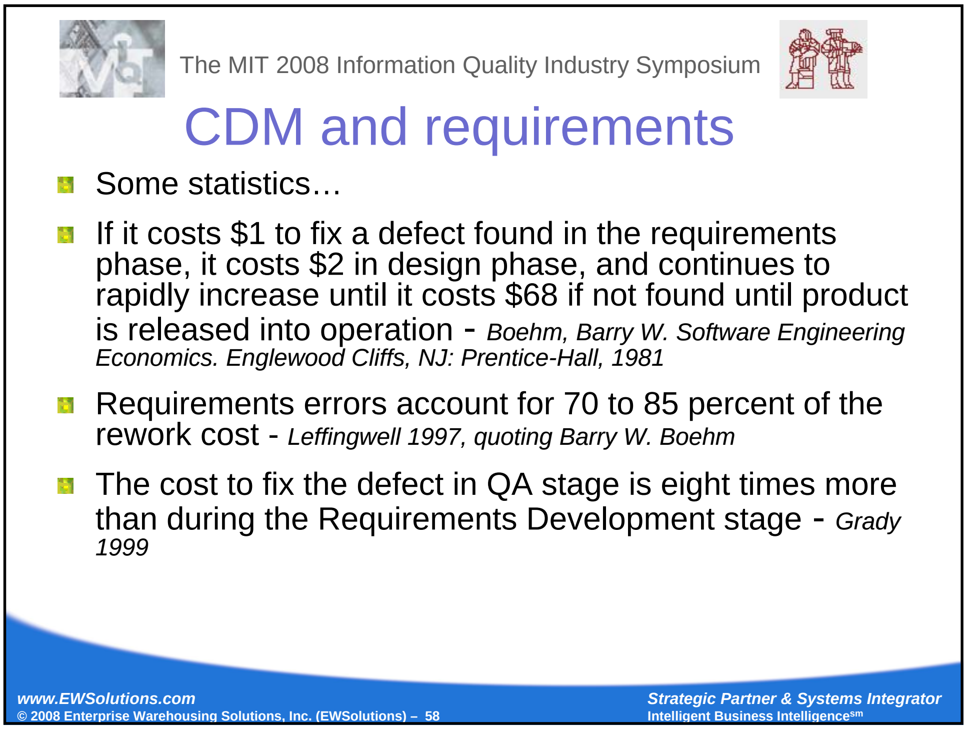
$ $ $ $ During application development

$ $ $$$$ During implementation

$ $ $$$ $$$$ During production
$ $ $$

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 57

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



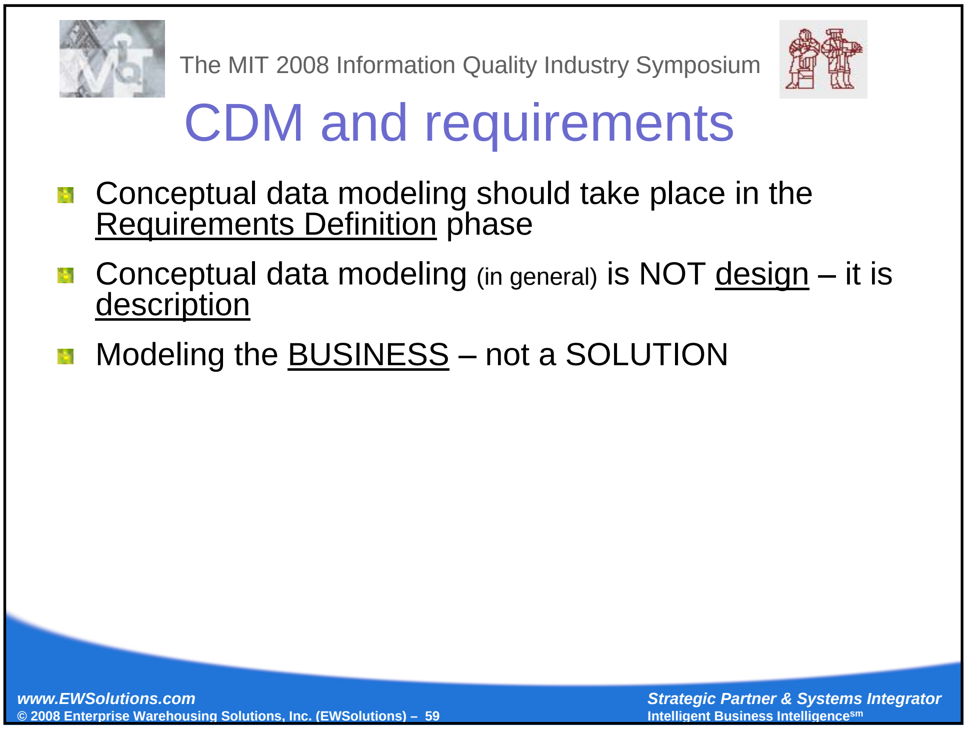
The MIT 2008 Information Quality Industry Symposium

CDM and requirements

- Some statistics…
- If it costs \$1 to fix a defect found in the requirements phase, it costs \$2 in design phase, and continues to rapidly increase until it costs \$68 if not found until product is released into operation - *Boehm, Barry W. Software Engineering Economics. Englewood Cliffs, NJ: Prentice-Hall, 1981*
- Requirements errors account for 70 to 85 percent of the rework cost - *Leffingwell 1997, quoting Barry W. Boehm*
- The cost to fix the defect in QA stage is eight times more than during the Requirements Development stage - *Grady 1999*

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 58

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

CDM and requirements

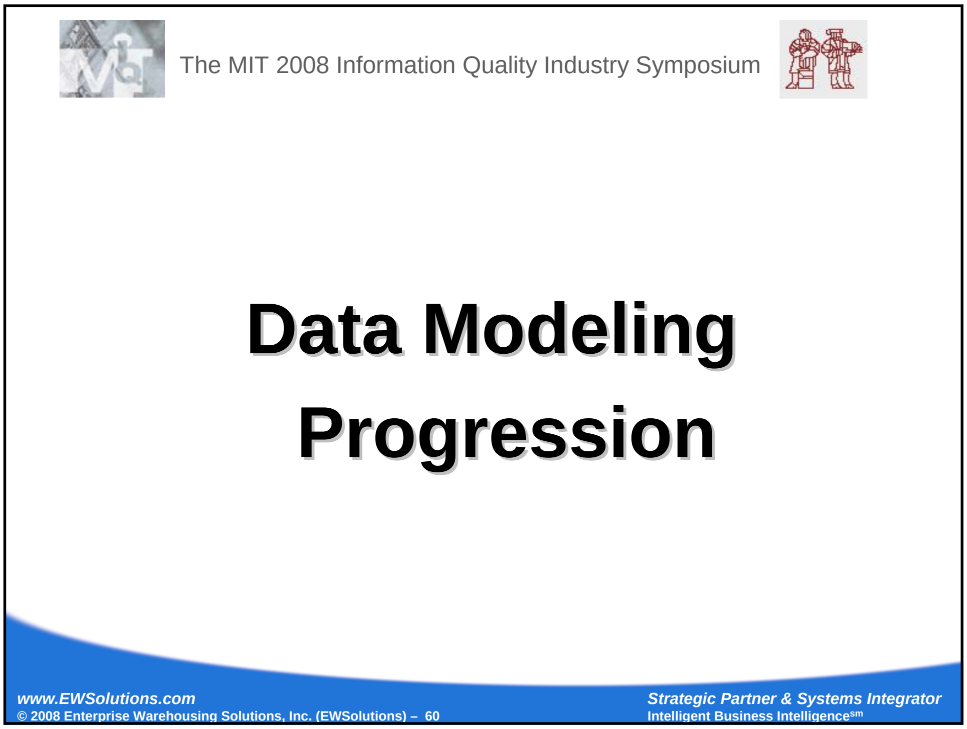
- Conceptual data modeling should take place in the Requirements Definition phase
- Conceptual data modeling (in general) is NOT design – it is description
- Modeling the BUSINESS – not a SOLUTION

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 59

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



The MIT 2008 Information Quality Industry Symposium

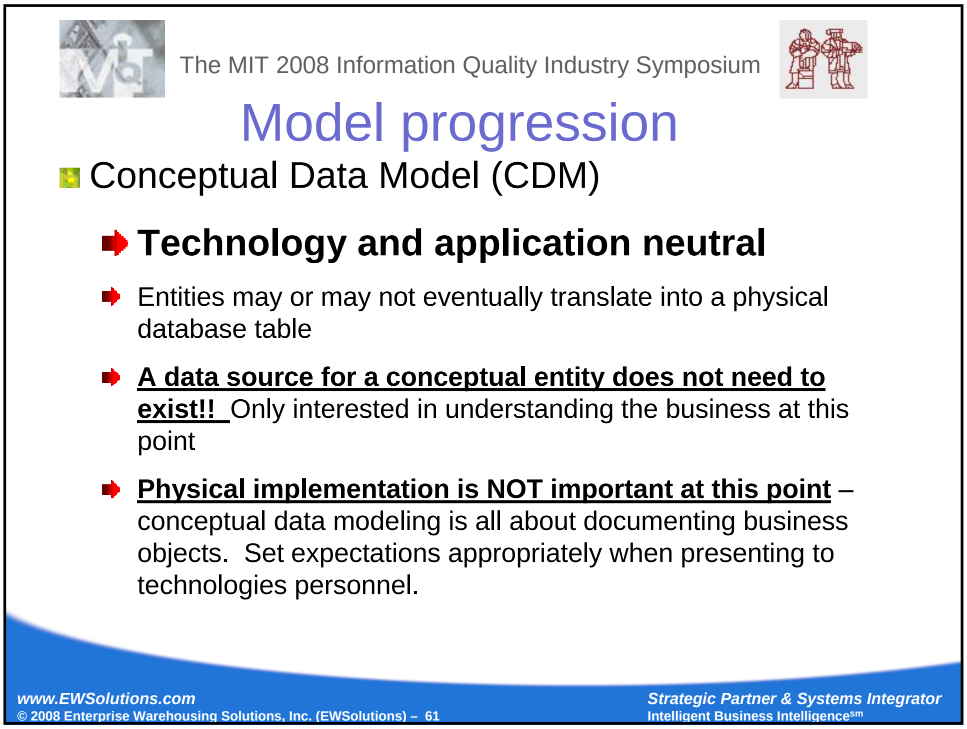
Data Modeling Progression

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 60

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^{sm}



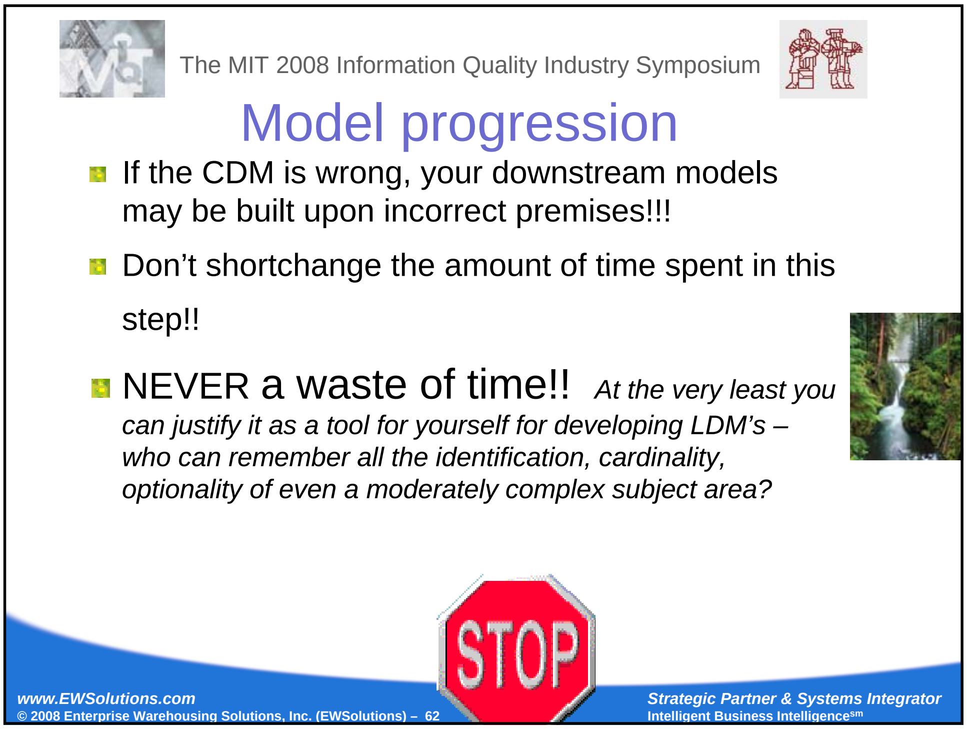
The MIT 2008 Information Quality Industry Symposium

Model progression

- Conceptual Data Model (CDM)
  - **Technology and application neutral**
  - Entities may or may not eventually translate into a physical database table
  - **A data source for a conceptual entity does not need to exist!!** Only interested in understanding the business at this point
  - **Physical implementation is NOT important at this point** – conceptual data modeling is all about documenting business objects. Set expectations appropriately when presenting to technologies personnel.

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 61

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Model progression

- If the CDM is wrong, your downstream models may be built upon incorrect premises!!!
- Don’t shortchange the amount of time spent in this step!!
- NEVER a waste of time!! *At the very least you can justify it as a tool for yourself for developing LDM’s – who can remember all the identification, cardinality, optionality of even a moderately complex subject area?*

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 62

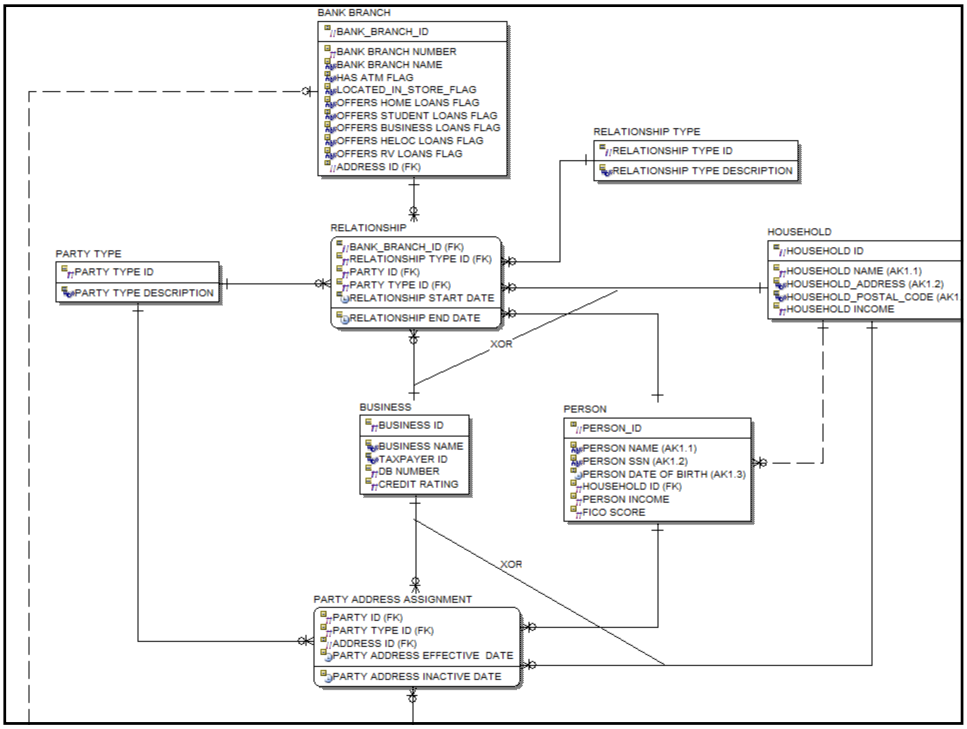
Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠

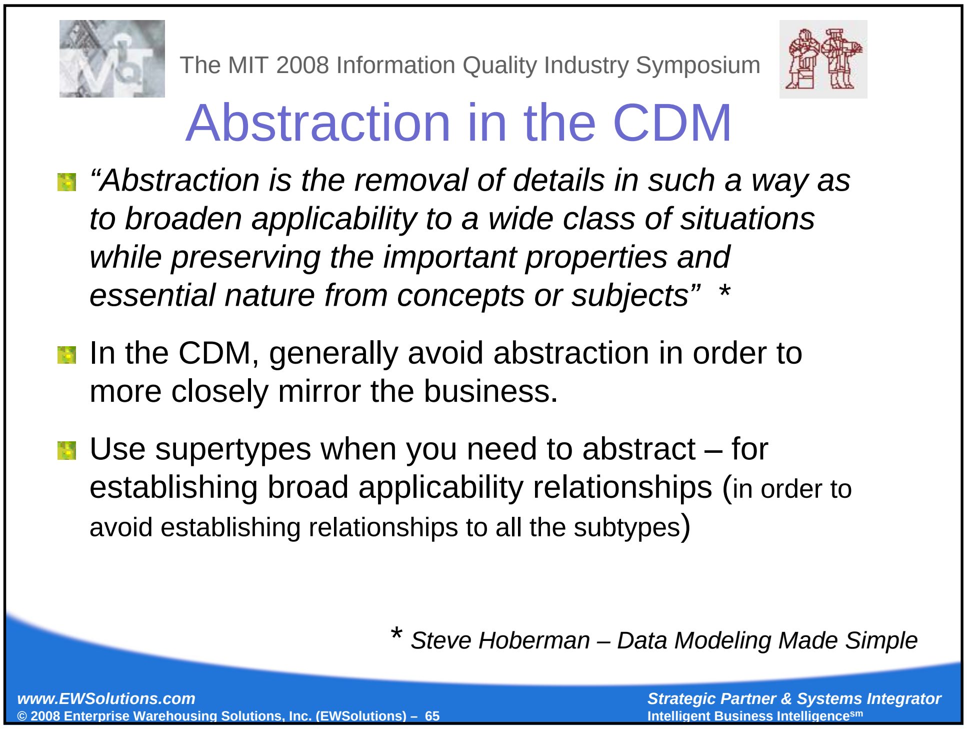
The MIT 2008 Information Quality Industry Symposium

Model progression

- Logical Data Model (LDM)
  - First step of <u>SOLUTION DATA DESIGN</u> (generally)
  - Fully/mostly attributizes a conceptual data model
  - Resolves many to many relationships (usually)
  - Resolves subtypes/supertypes (usually)
  - May introduce abstraction (generalize entities, attributes, relationships) – more later
  - Formalize primary keys

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 63



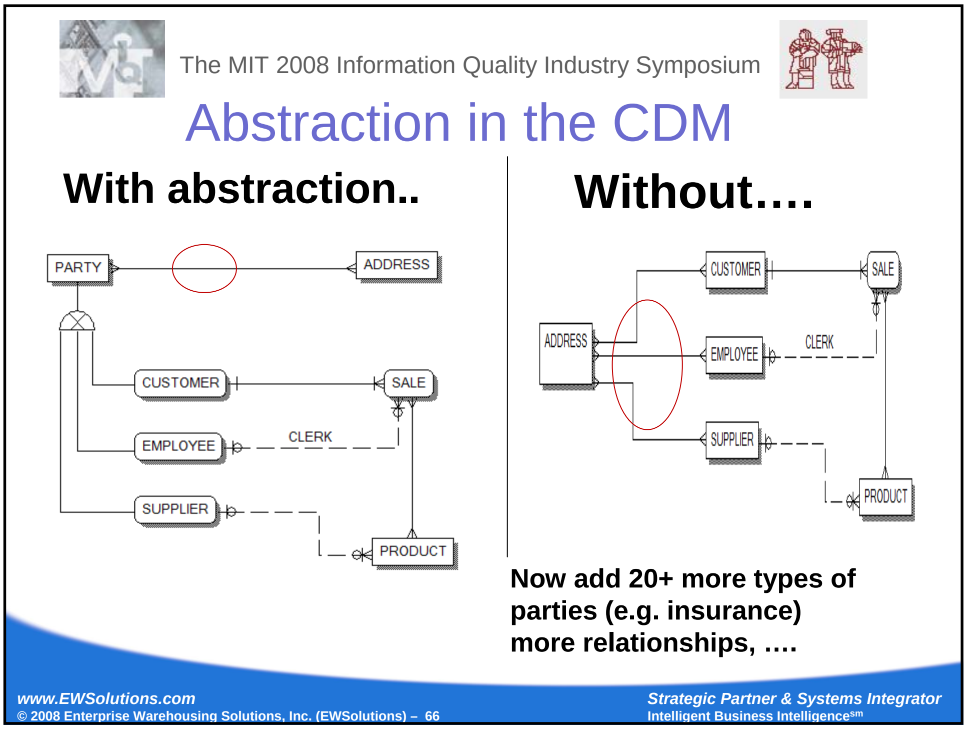


The MIT 2008 Information Quality Industry Symposium

Abstraction in the CDM

- *"Abstraction is the removal of details in such a way as to broaden applicability to a wide class of situations while preserving the important properties and essential nature from concepts or subjects"* *
- In the CDM, generally avoid abstraction in order to more closely mirror the business.
- Use supertypes when you need to abstract – for establishing broad applicability relationships (in order to avoid establishing relationships to all the subtypes)

* *Steve Hoberman – Data Modeling Made Simple*



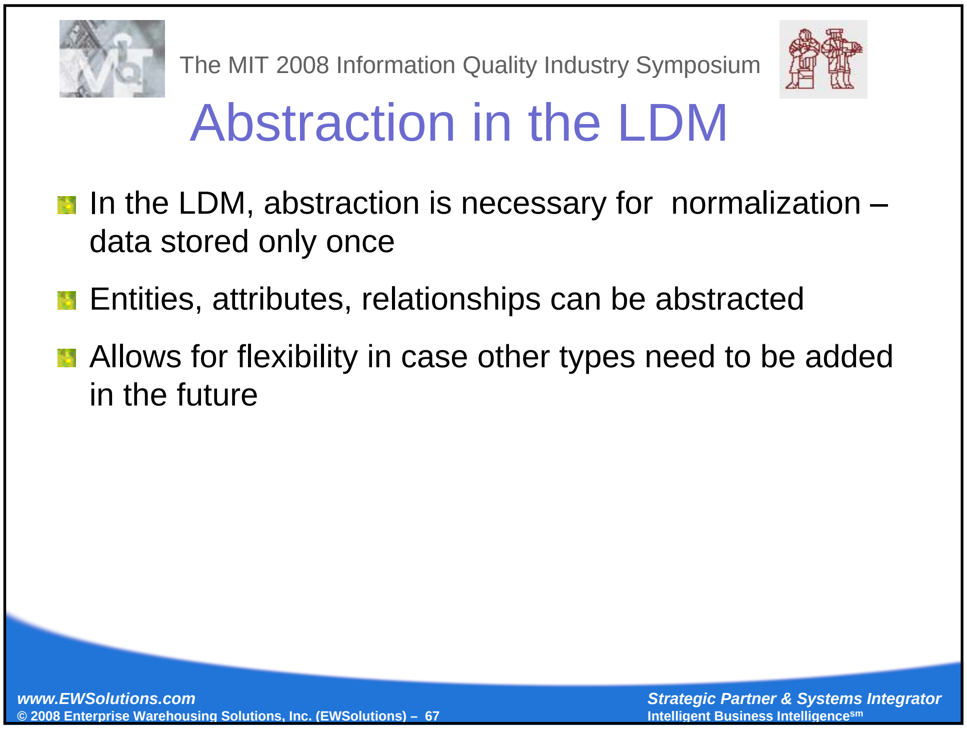
The MIT 2008 Information Quality Industry Symposium

Abstraction in the CDM

With abstraction..

Without….

Now add 20+ more types of parties (e.g. insurance) more relationships, ….



The MIT 2008 Information Quality Industry Symposium

Abstraction in the LDM

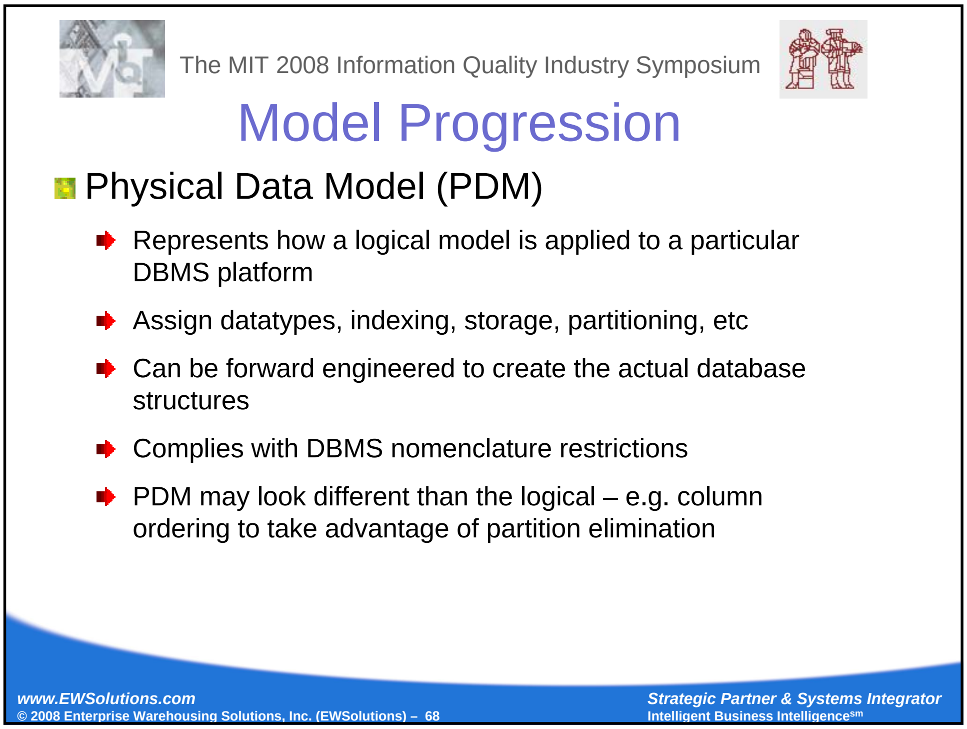
- In the LDM, abstraction is necessary for normalization – data stored only once
- Entities, attributes, relationships can be abstracted
- Allows for flexibility in case other types need to be added in the future

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 67

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^sm



The MIT 2008 Information Quality Industry Symposium

Model Progression

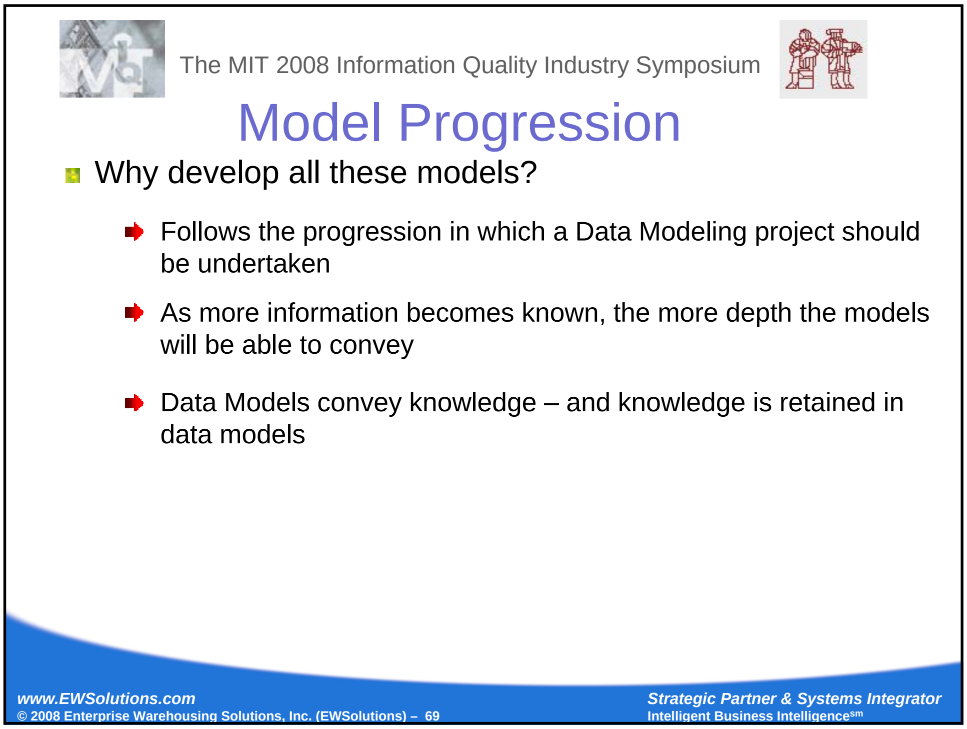
- Physical Data Model (PDM)
  - Represents how a logical model is applied to a particular DBMS platform
  - Assign datatypes, indexing, storage, partitioning, etc
  - Can be forward engineered to create the actual database structures
  - Complies with DBMS nomenclature restrictions
  - PDM may look different than the logical – e.g. column ordering to take advantage of partition elimination

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 68

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^sm



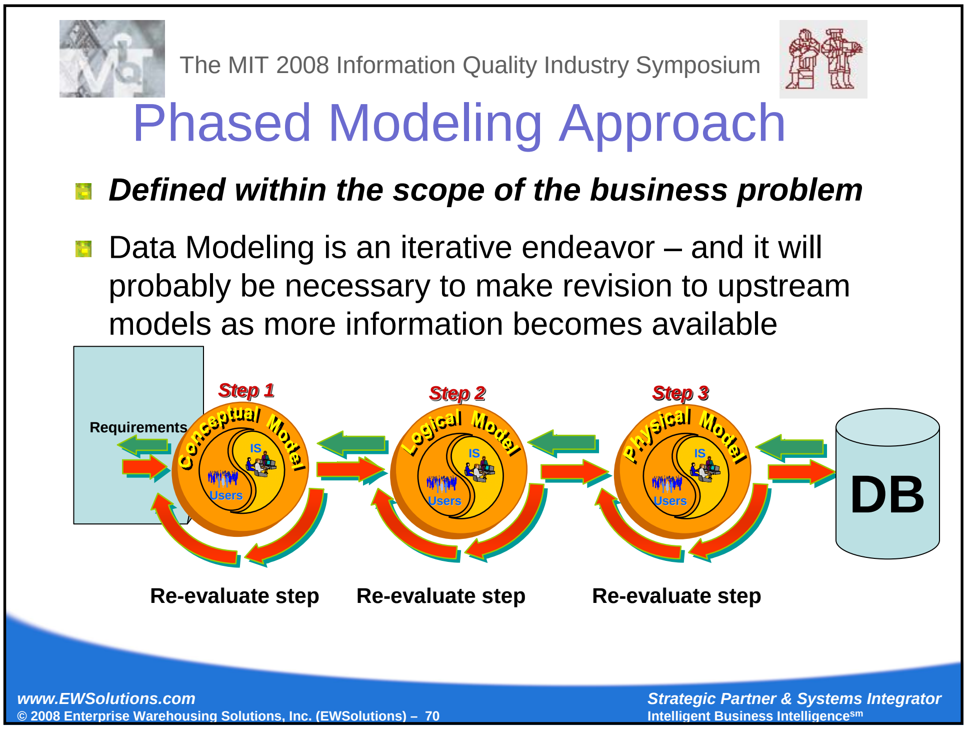
The MIT 2008 Information Quality Industry Symposium

Model Progression

- Why develop all these models?
  - Follows the progression in which a Data Modeling project should be undertaken
  - As more information becomes known, the more depth the models will be able to convey
  - Data Models convey knowledge – and knowledge is retained in data models

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 69

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}



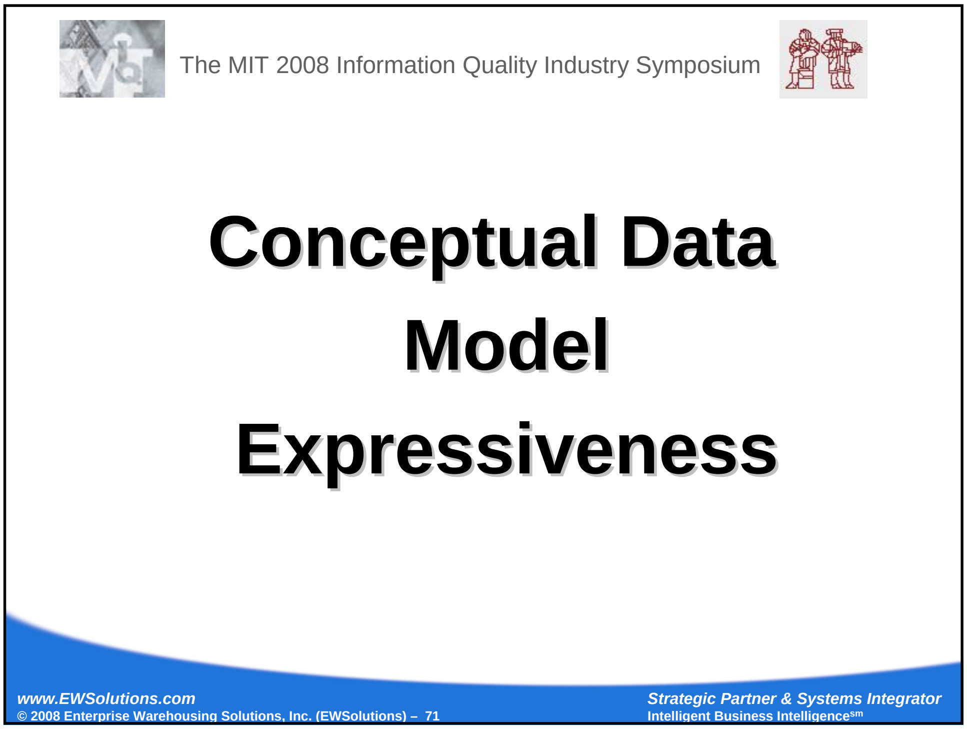
The MIT 2008 Information Quality Industry Symposium

Phased Modeling Approach

- ***Defined within the scope of the business problem***
- Data Modeling is an iterative endeavor – and it will probably be necessary to make revision to upstream models as more information becomes available

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 70

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}

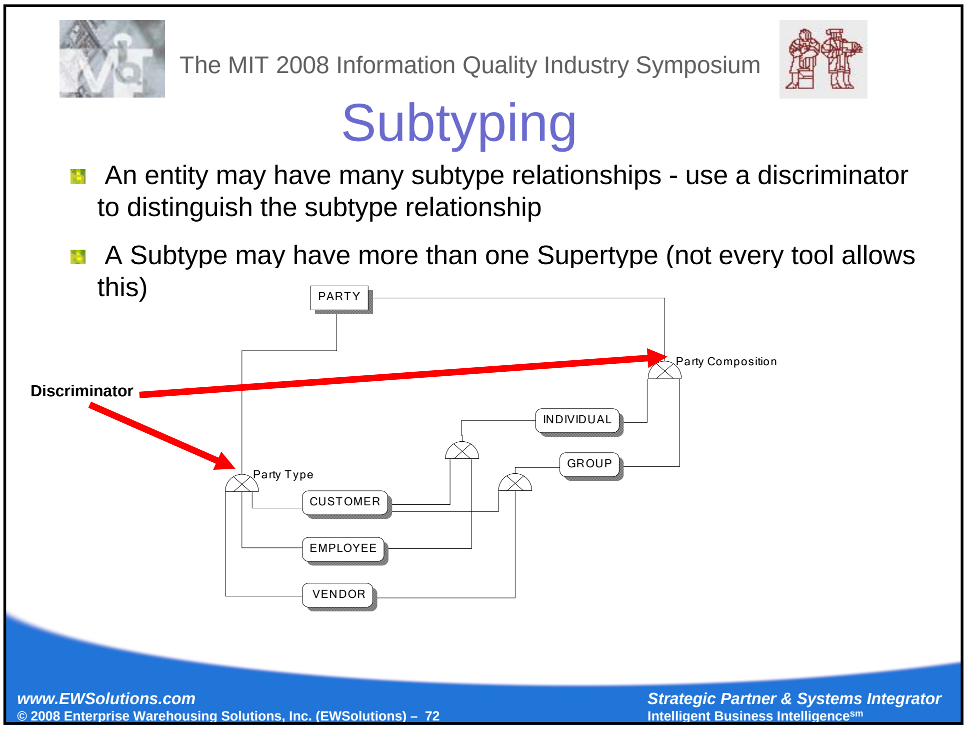


The MIT 2008 Information Quality Industry Symposium

Conceptual Data Model Expressiveness

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 71

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



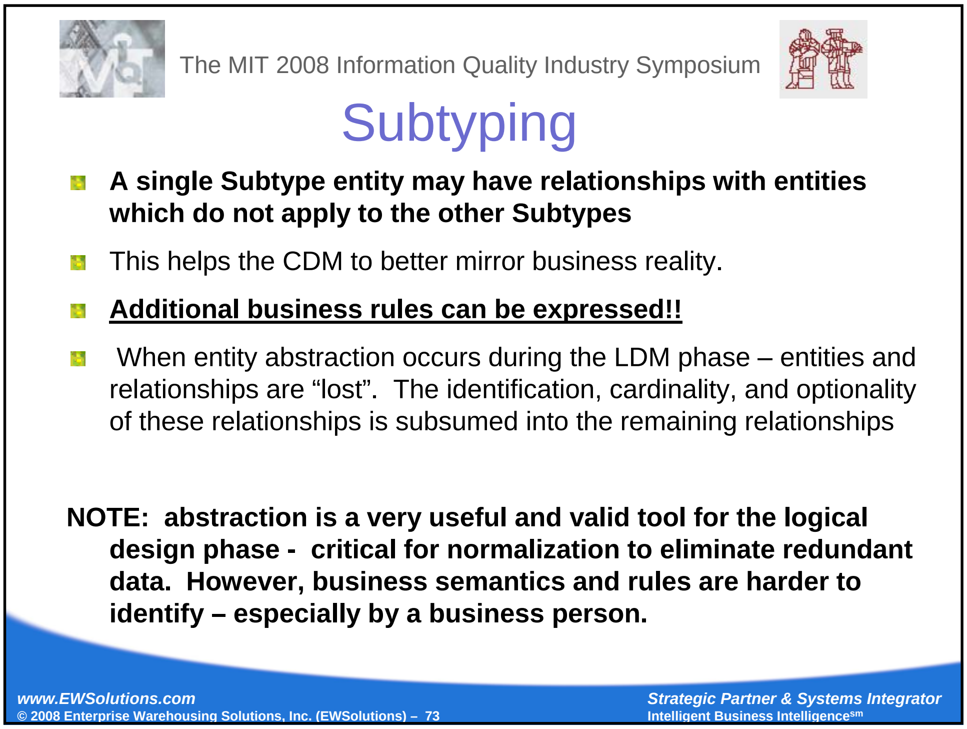
The MIT 2008 Information Quality Industry Symposium

Subtyping

- An entity may have many subtype relationships - use a discriminator to distinguish the subtype relationship
- A Subtype may have more than one Supertype (not every tool allows this)

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 72

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

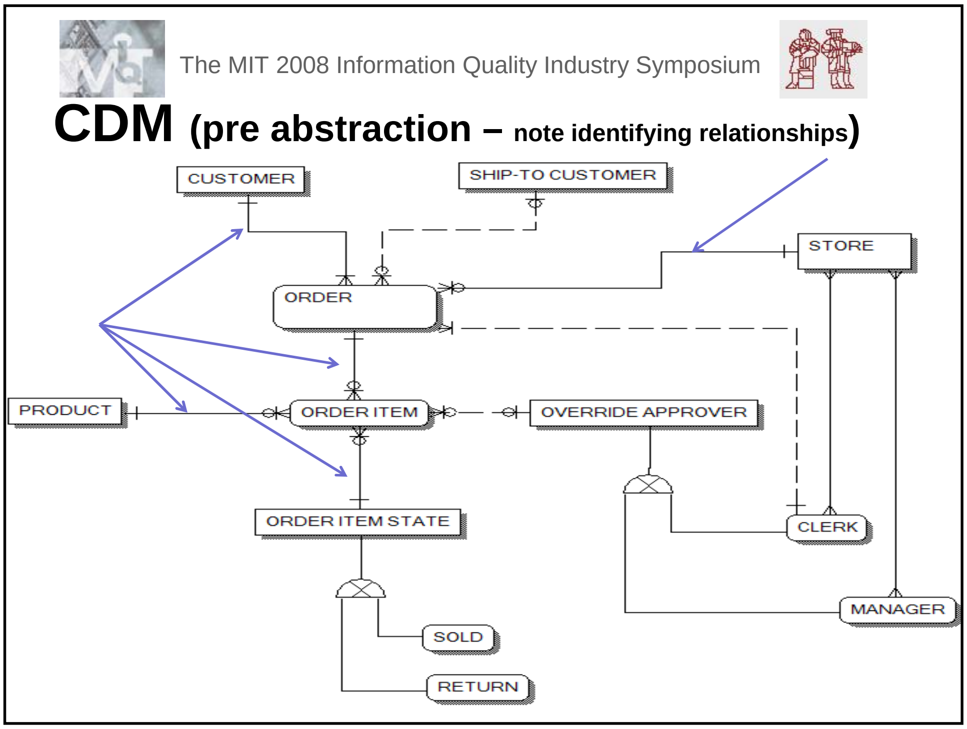
Subtyping

- **A single Subtype entity may have relationships with entities which do not apply to the other Subtypes**
- This helps the CDM to better mirror business reality.
- **<u>Additional business rules can be expressed!!</u>**
- When entity abstraction occurs during the LDM phase – entities and relationships are "lost". The identification, cardinality, and optionality of these relationships is subsumed into the remaining relationships

NOTE: abstraction is a very useful and valid tool for the logical design phase - critical for normalization to eliminate redundant data. However, business semantics and rules are harder to identify – especially by a business person.

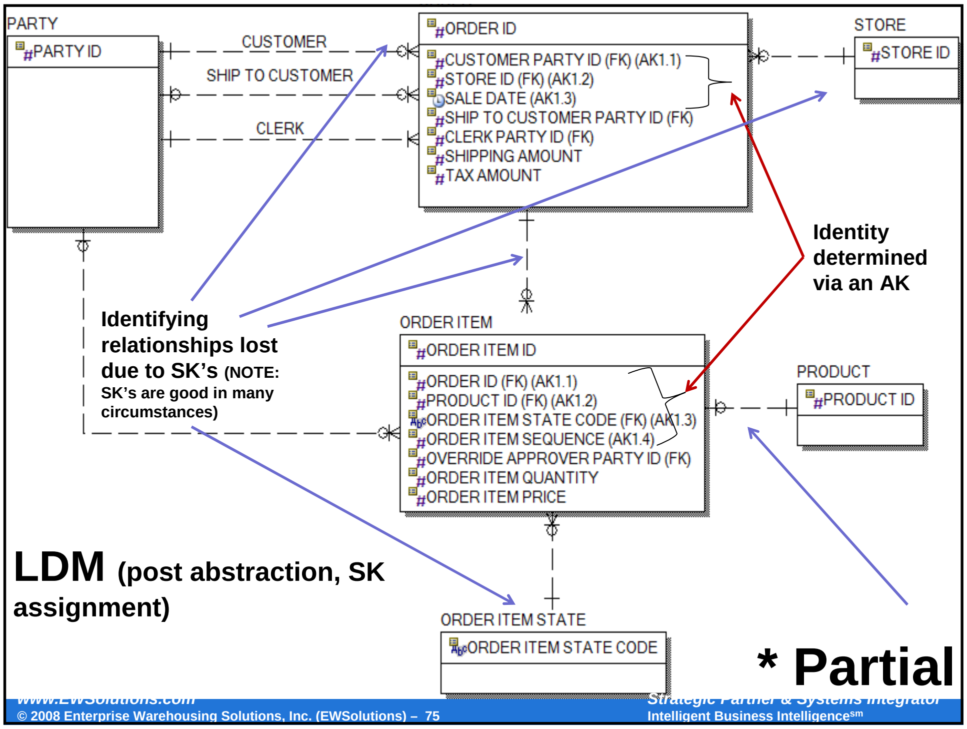
www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 73

Strategic Partner & Systems Integrator
Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

CDM (pre abstraction – note identifying relationships)



PARTY

PARTY ID

CUSTOMER

SHIP TO CUSTOMER

CLERK

ORDER ID

CUSTOMER PARTY ID (FK) (AK1.1)
STORE ID (FK) (AK1.2)
SALE DATE (AK1.3)
SHIP TO CUSTOMER PARTY ID (FK)
CLERK PARTY ID (FK)
SHIPPING AMOUNT
TAX AMOUNT

STORE

STORE ID

Identity determined via an AK

Identifying relationships lost due to SK's (NOTE: SK's are good in many circumstances)

ORDER ITEM

ORDER ITEM ID

ORDER ID (FK) (AK1.1)
PRODUCT ID (FK) (AK1.2)
ORDER ITEM STATE CODE (FK) (AK1.3)
ORDER ITEM SEQUENCE (AK1.4)
OVERRIDE APPROVER PARTY ID (FK)
ORDER ITEM QUANTITY
ORDER ITEM PRICE

PRODUCT

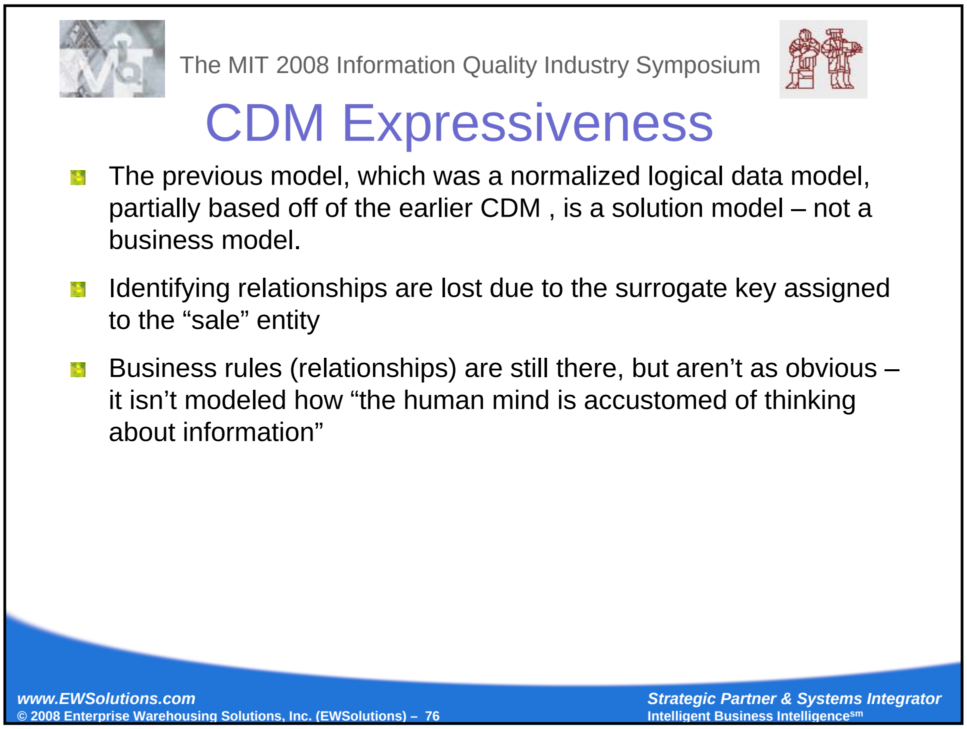
PRODUCT ID

LDM (post abstraction, SK assignment)

ORDER ITEM STATE

ORDER ITEM STATE CODE

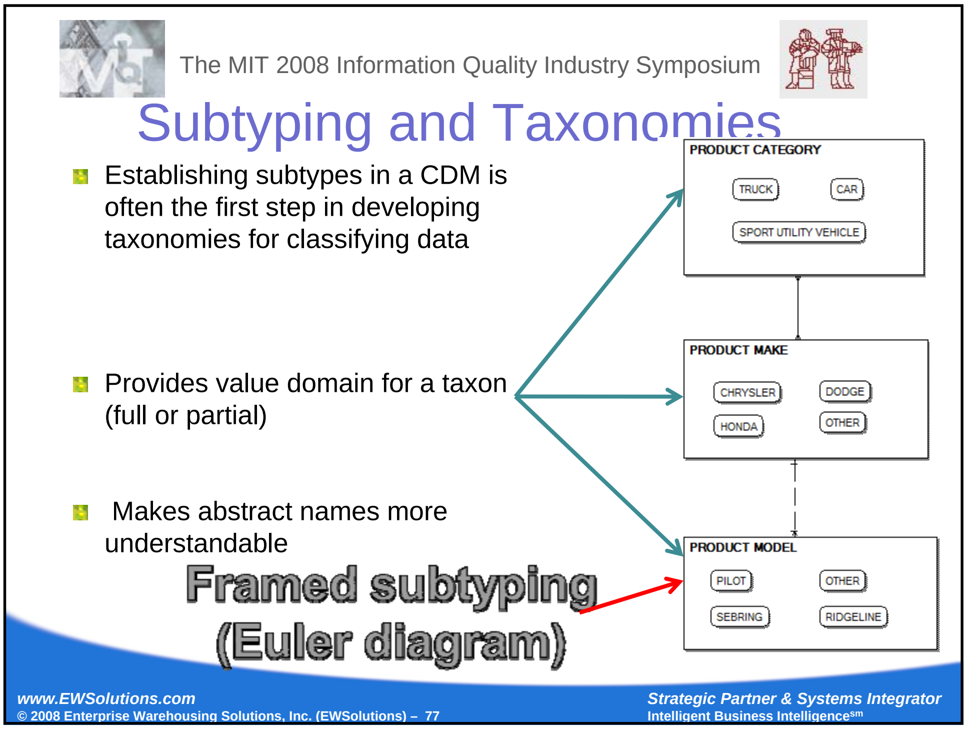
*** Partial**



The MIT 2008 Information Quality Industry Symposium

CDM Expressiveness

- The previous model, which was a normalized logical data model, partially based off of the earlier CDM , is a solution model – not a business model.
- Identifying relationships are lost due to the surrogate key assigned to the "sale" entity
- Business rules (relationships) are still there, but aren't as obvious – it isn't modeled how "the human mind is accustomed of thinking about information"



The MIT 2008 Information Quality Industry Symposium

Subtyping and Taxonomies

- Establishing subtypes in a CDM is often the first step in developing taxonomies for classifying data
- Provides value domain for a taxon (full or partial)
- Makes abstract names more understandable

Framed subtyping (Euler diagram)

PRODUCT CATEGORY

TRUCK

CAR

SPORT UTILITY VEHICLE

PRODUCT MAKE

CHRYSLER

DODGE

HONDA

OTHER

PRODUCT MODEL

PILOT

OTHER

SEBRING

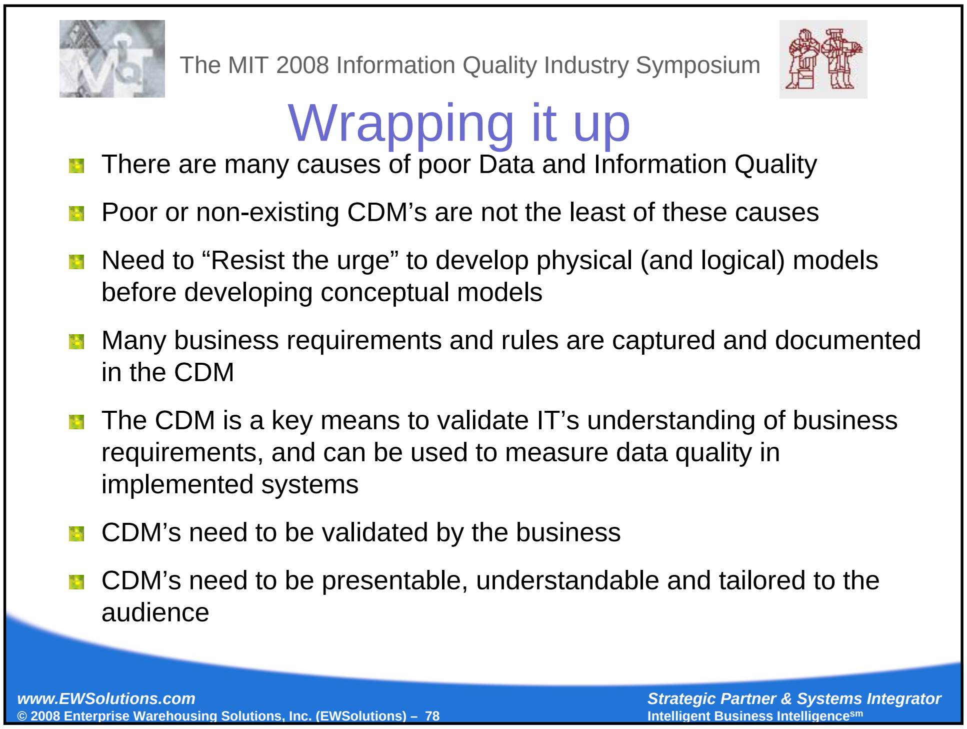
RIDGELINE

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 77

Strategic Partner & Systems Integrator

Intelligent Business Intelligence℠



The MIT 2008 Information Quality Industry Symposium

Wrapping it up

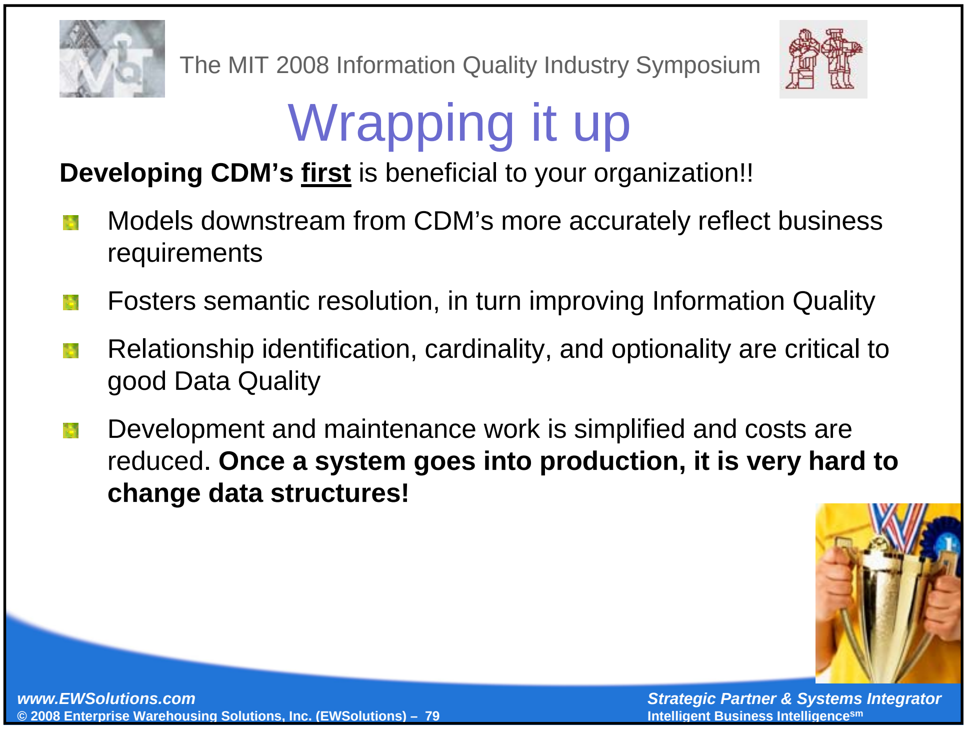
- There are many causes of poor Data and Information Quality
- Poor or non-existing CDM's are not the least of these causes
- Need to "Resist the urge" to develop physical (and logical) models before developing conceptual models
- Many business requirements and rules are captured and documented in the CDM
- The CDM is a key means to validate IT's understanding of business requirements, and can be used to measure data quality in implemented systems
- CDM's need to be validated by the business
- CDM's need to be presentable, understandable and tailored to the audience

www.EWSolutions.com

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 78

Strategic Partner & Systems Integrator

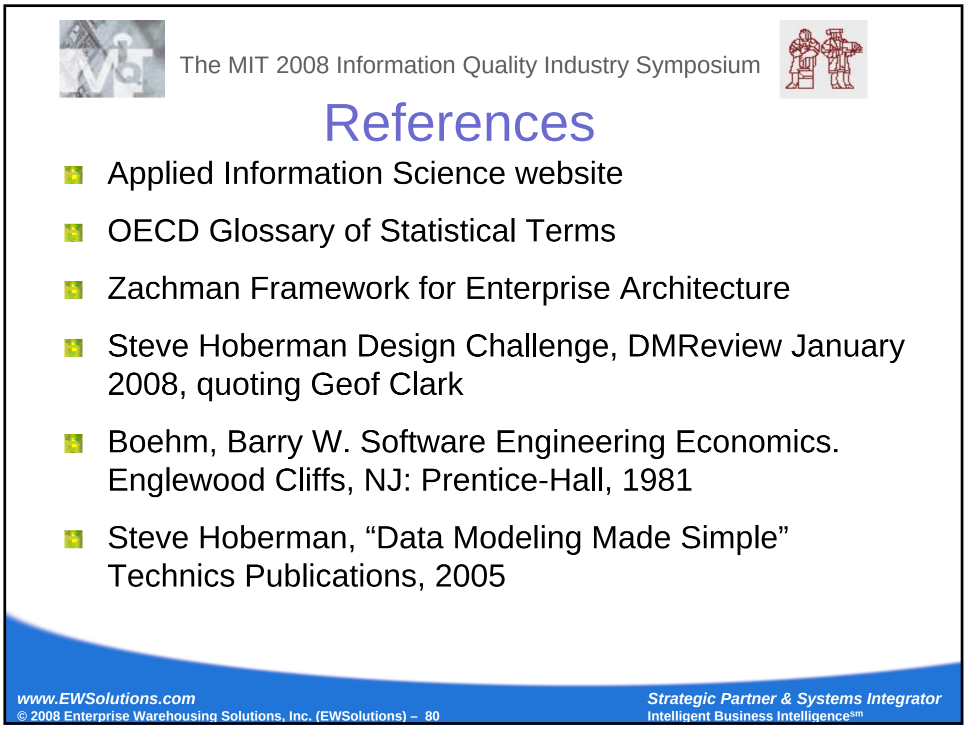
Intelligent Business Intelligence℠



Wrapping it up

Developing CDM's <u>first</u> is beneficial to your organization!!

- Models downstream from CDM's more accurately reflect business requirements
- Fosters semantic resolution, in turn improving Information Quality
- Relationship identification, cardinality, and optionality are critical to good Data Quality
- Development and maintenance work is simplified and costs are reduced. **Once a system goes into production, it is very hard to change data structures!**



References

- Applied Information Science website
- OECD Glossary of Statistical Terms
- Zachman Framework for Enterprise Architecture
- Steve Hoberman Design Challenge, DMReview January 2008, quoting Geof Clark
- Boehm, Barry W. Software Engineering Economics. Englewood Cliffs, NJ: Prentice-Hall, 1981
- Steve Hoberman, "Data Modeling Made Simple" Technics Publications, 2005

The MIT 2008 Information Quality Industry Symposium

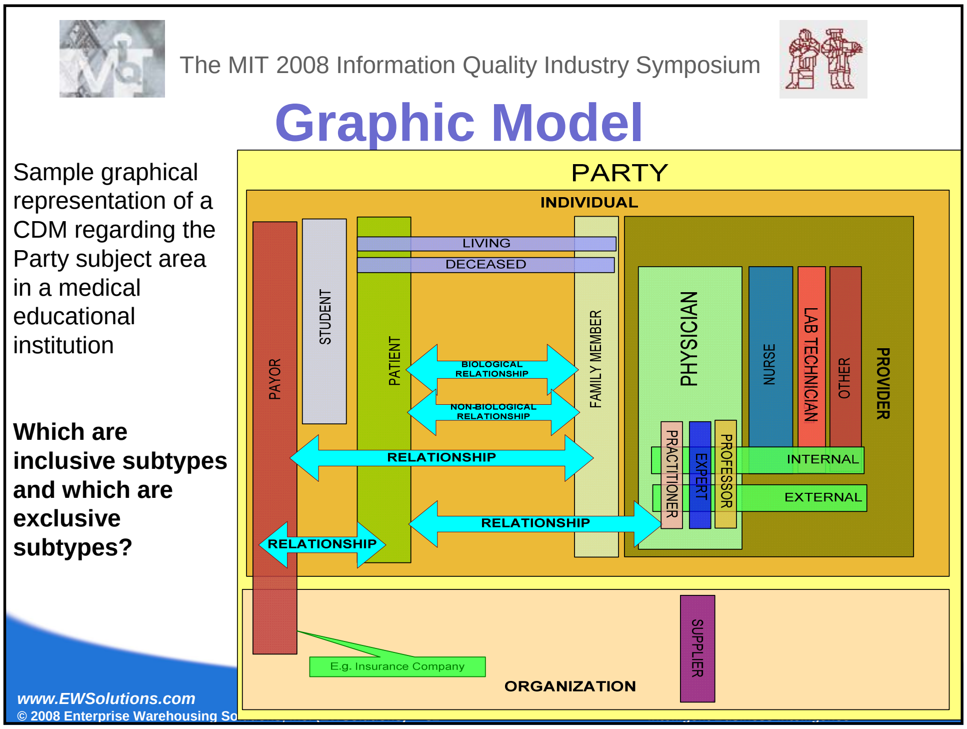
Questions?

www.EWSolutions.com

Strategic Partner & Systems Integrator

Intelligent Business Intelligence^sm

© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 81



The MIT 2008 Information Quality Industry Symposium

Graphic Model

Sample graphical representation of a CDM regarding the Party subject area in a medical educational institution

Which are inclusive subtypes and which are exclusive subtypes?

PARTY

INDIVIDUAL

PAYOR

STUDENT

PATIENT

LIVING

DECEASED

BIOLOGICAL RELATIONSHIP

NON-BIOLOGICAL RELATIONSHIP

RELATIONSHIP

RELATIONSHIP

RELATIONSHIP

FAMILY MEMBER

PHYSICIAN

PRACTITIONER

EXPERT

PROFESSOR

NURSE

LAB TECHNICIAN

OTHER

PROVIDER

INTERNAL

EXTERNAL

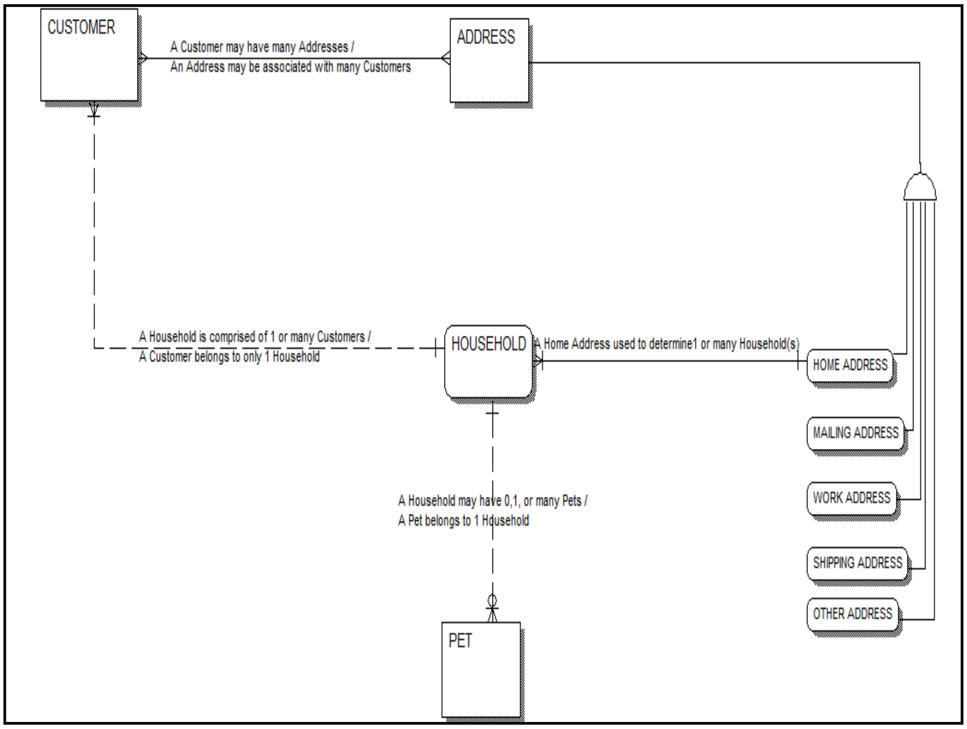
E.g. Insurance Company

SUPPLIER

ORGANIZATION

www.EWSolutions.com

© 2008 Enterprise Warehousing So


CUSTOMER
A Customer may have many Addresses /
An Address may be associated with many Customers
ADDRESS
A Household is comprised of 1 or many Customers /
A Customer belongs to only 1 Household
HOUSEHOLD
A Home Address used to determine1 or many Household(s)
HOME ADDRESS
MAILING ADDRESS
WORK ADDRESS
SHIPPING ADDRESS
OTHER ADDRESS
A Household may have 0,1, or many Pets /
A Pet belongs to 1 Household
PET

The MIT 2008 Information Quality Industry Symposium

EWSolutions, Inc.
15 Spinning Wheel Road,
Suite 330
Hinsdale, IL 60521
Office 630.920.0005
Fax 630.920.0008

http://www.EWSolutions.com

www.EWSolutions.com
© 2008 Enterprise Warehousing Solutions, Inc. (EWSolutions) – 84

Strategic Partner & Systems Integrator
Intelligent Business Intelligence^{sm}