Data Modeling is Fundamental

Peter Aiken, Ph.D.

Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Founder, Data Blueprint (datablueprint.com)
- DAMA International (dama.org)
- 10 books and dozens of articles
- Experienced w/ 500+ data management practices worldwide
- Multi-year immersions
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - …

UNLOCKING BUSINESS VALUE

Data Modeling Fundamentals

- **Data Management Contextual Overview**
- Motivation
  - of systems/components
  - Data is not well understood
- Why data modeling & what is it?
  - Model represents our understanding of the
  - Fundamental, foundational system characteristics
  - Shared between system and human
- Fundamentals
  - The power of the purpose statement
  - Understanding data centric thinking
  - Data modeling compliments other architecture/ engineering techniques, as well as
  - Challenges beyond data modeling
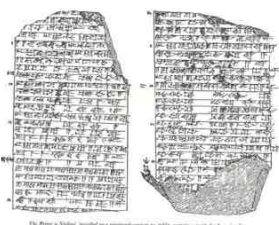- Take Aways, References & Q&A

Mark your calendar for our monthly #DataEd webinars

DATAVERSITY

What is the world's oldest profession?

The Oldest Profession

ACCOUNTANT

- 8,000+ years
- formalize practices
- GAAP

GAAP
Generally
Accepted
Accounting
Principles
Rules · Standards · Procedures

It is appropriate that we (data professionals) acknowledge that we are currently not as mature a discipline as we would like to be but it is <u>not</u> okay for our discipline to remain in its current state of maturity
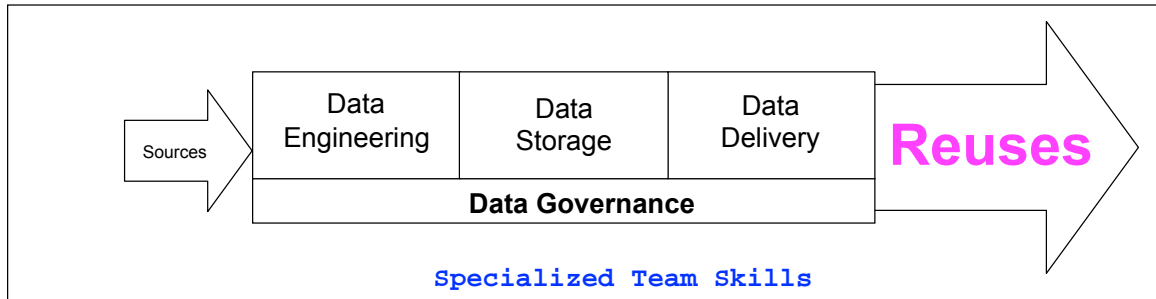
Augusta Ada King
Countess of Lovelace
(1815-52)

What is data management?

Sources →	Data Engineering	Data Storage	Data Delivery	**Reuses** →
	Data Governance			

Specialized Team Skills

Understanding the current and future data needs of an enterprise and making that data effective and efficient in supporting business activities

Aiken, P, Allen, M. D., Parker, B., Mattia, A.,
"Measuring Data Management's Maturity:
A Community's Self-Assessment"
IEEE Computer (research feature April 2007)

Data management practices connect data sources and uses in an organized and efficient manner
- Engineering
- Storage
- Delivery
- Governance

When executed, engineering, storage, and delivery implement governance
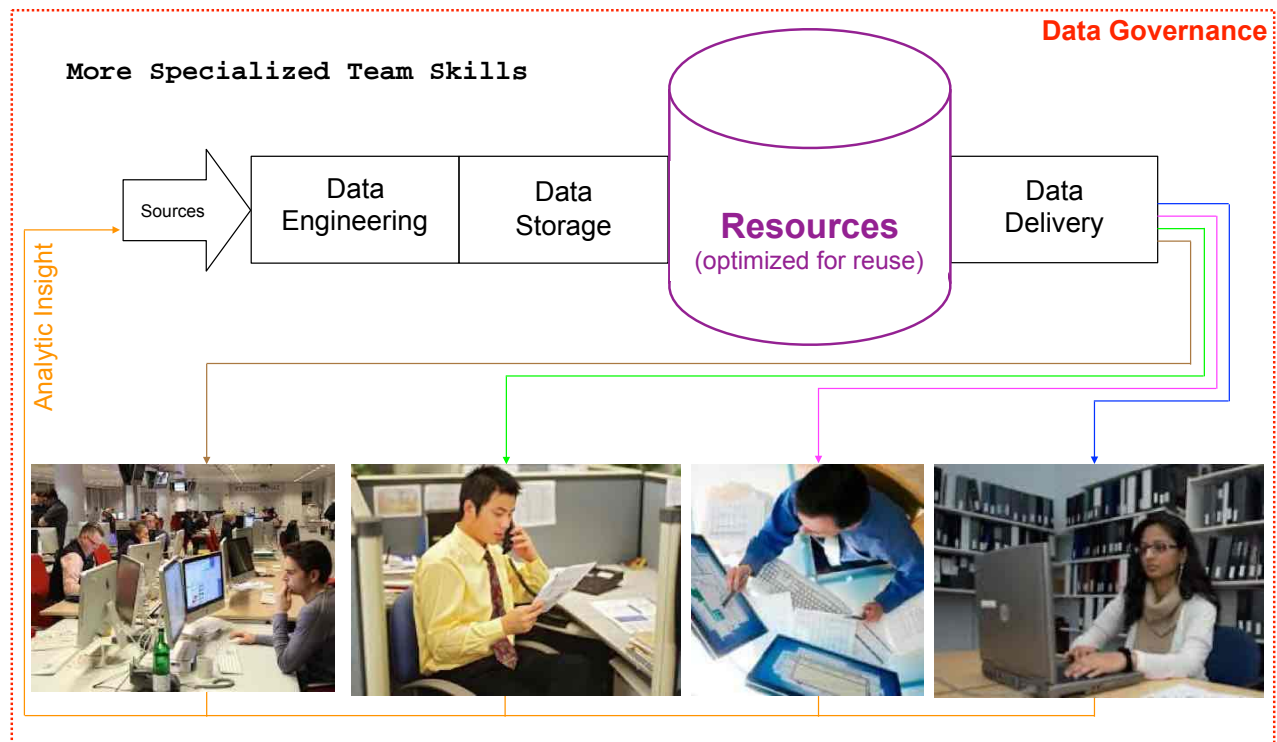
Note: does not well-depict data reuse

What is data management?

Data Governance

More Specialized Team Skills

Analytic Insight

Sources →	Data Engineering	Data Storage	**Resources** (optimized for reuse)	Data Delivery

Data Management Practices Hierarchy

You can accomplish Advanced Data Practices without becoming proficient in the Foundational Data Management Practices however this will:

- Take longer
- Cost more
- Deliver less
- Present greater risk

(with thanks to Tom DeMarco)

Technologies

Advanced Data Practices
- MDM
- Mining
- Big Data
- Analytics
- Warehousing
- SOA

Foundational Data Management Practices

Capabilities

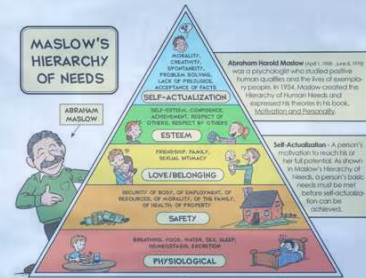Data Governance   Data Quality
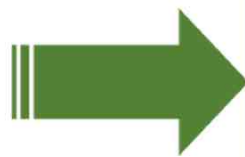
Data Management Strategy

Data Platform/Architecture   Data Operations

MASLOW'S HIERARCHY OF NEEDS

ABRAHAM MASLOW

SELF-ACTUALIZATION

ESTEEM

LOVE/BELONGING

SAFETY

PHYSIOLOGICAL

Recent Technology Realization

AWESOME THING

TRUE WITHOUT BLOCKCHAIN
TRUE WITH BLOCKCHAIN

BLOCKCHAIN (AWESOME)

Garbage In→Garbage Out!

QI➜QO!

Quality Data → Data Warehouse | Business Intelligence | Block Chain | Machine Learning | Data Governance | Analytics | MDM | AI Technology → Garbage Results

UNLOCKING BUSINESS VALUE

Copyright 2019 by Data Blueprint Slide # 13

Quality In ➜ Quality Out!

Quality Data → Data Warehouse | Business Intelligence | Block Chain | Machine Learning | Data Governance | Analytics | MDM | AI Technology → Good Results

UNLOCKING BUSINESS VALUE

Copyright 2019 by Data Blueprint Slide # 14
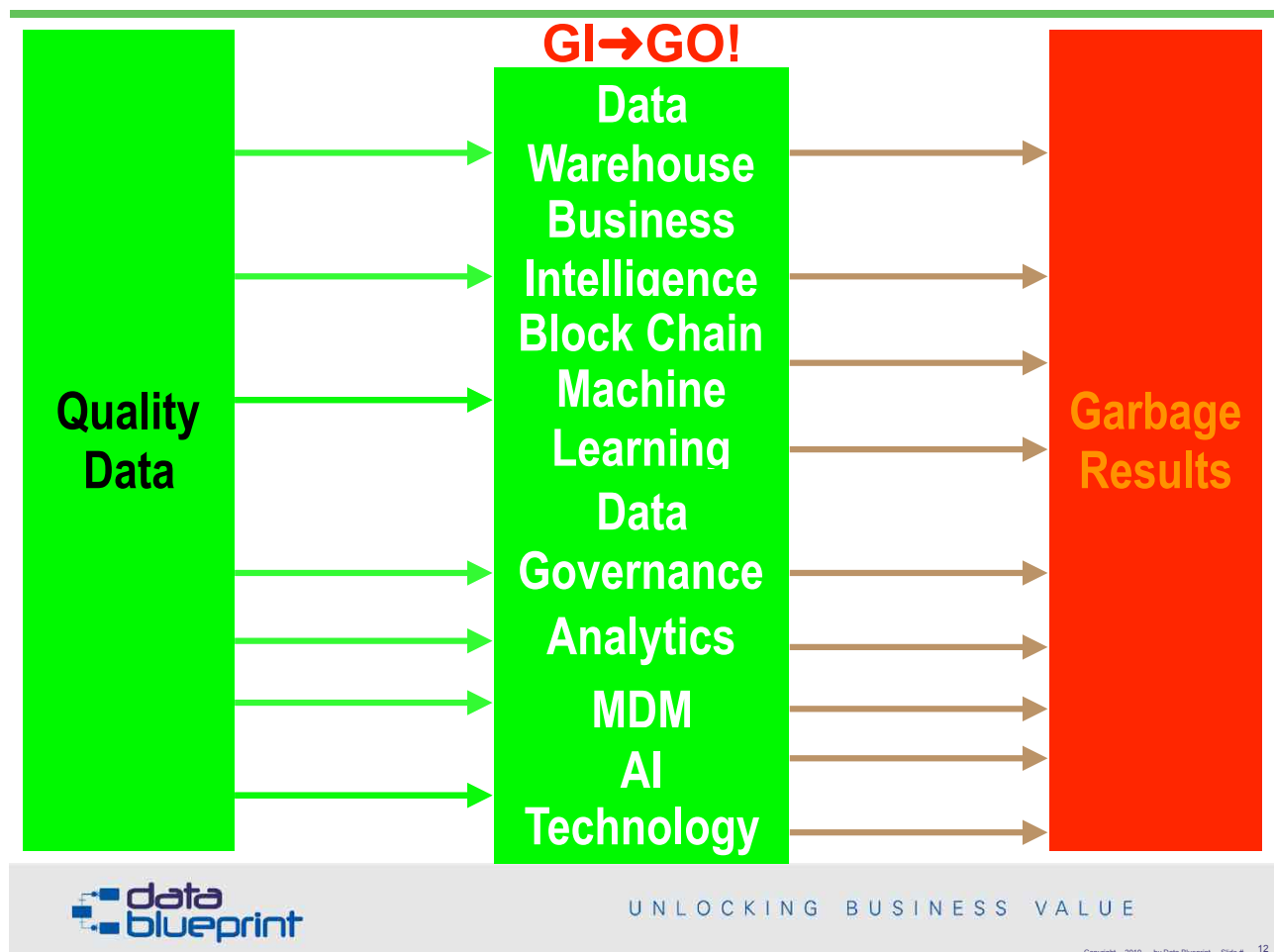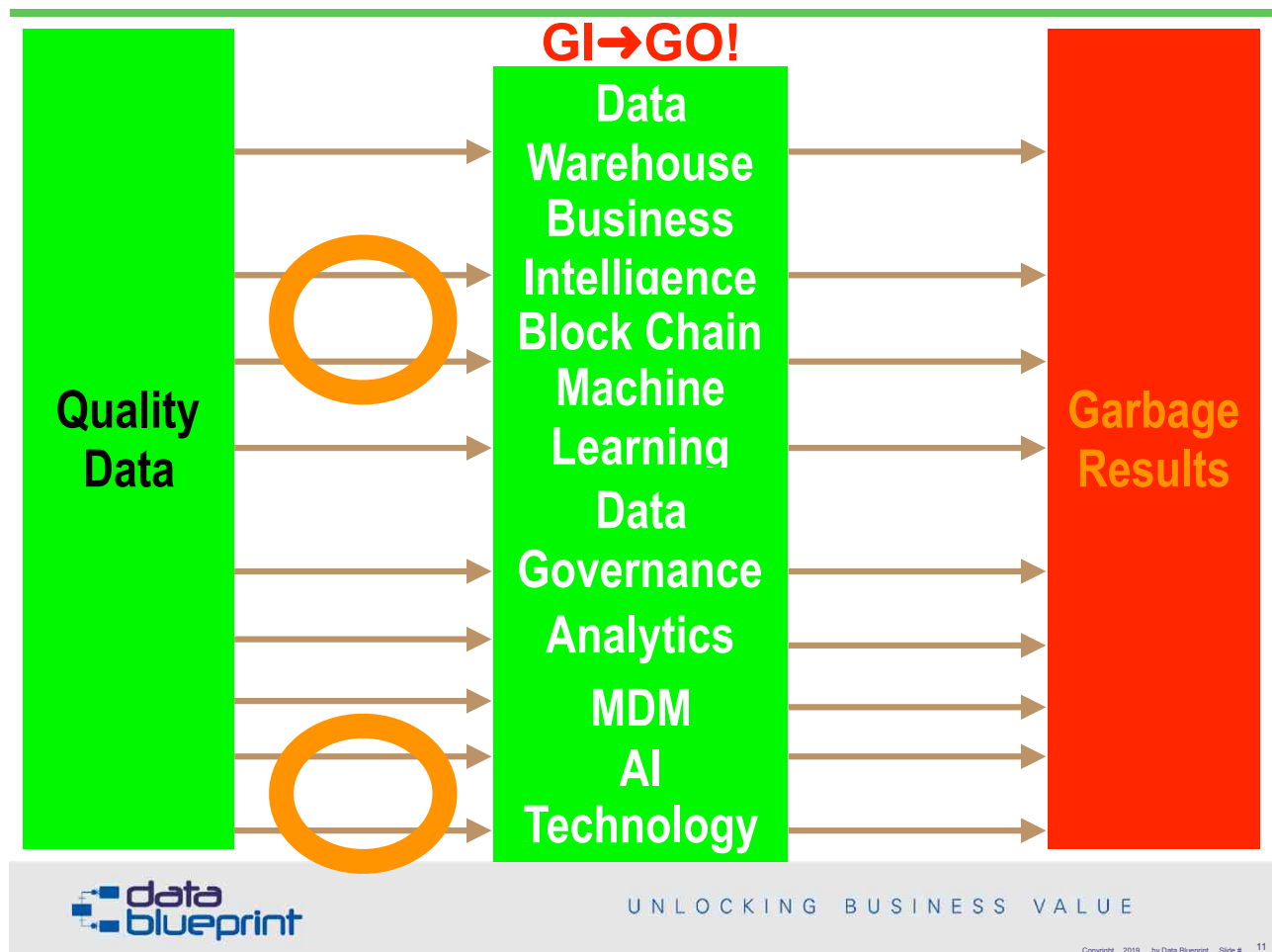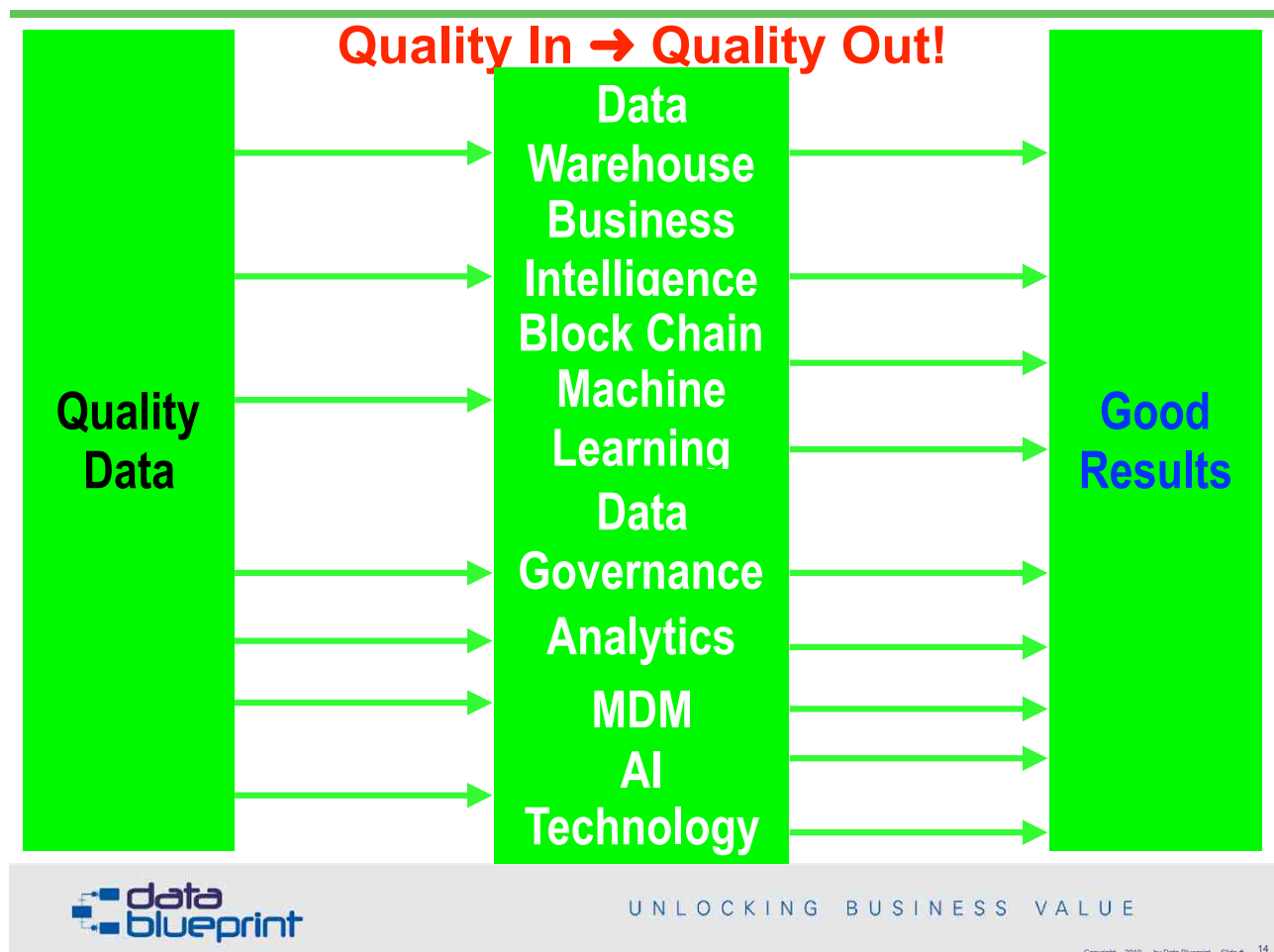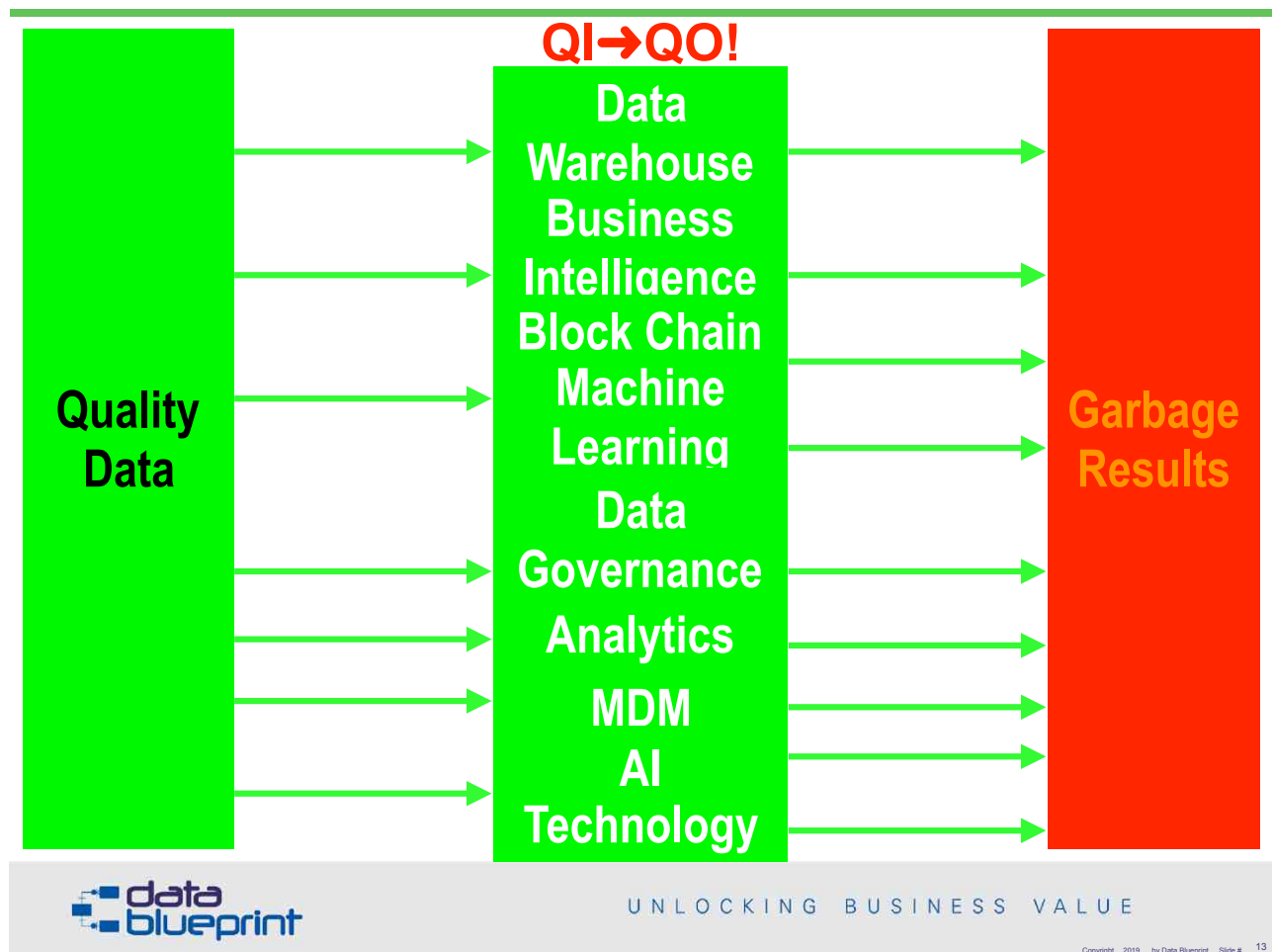
Slide 15

Data Management Body of Knowledge
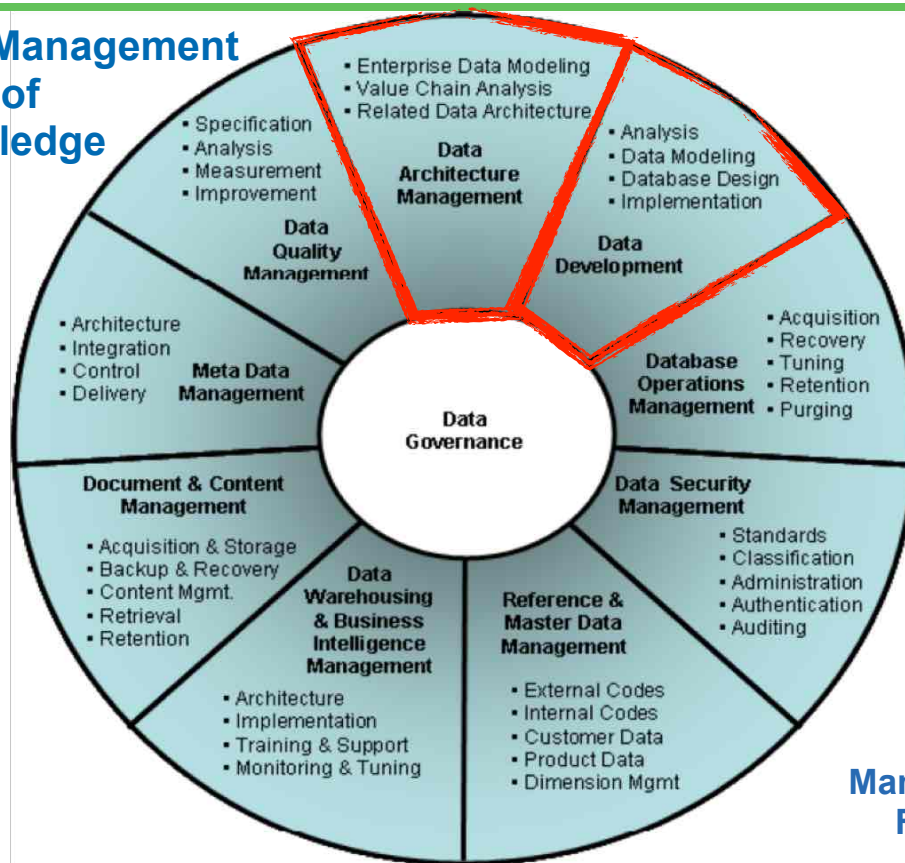
Data Management Functions

Data Architecture Management
- Enterprise Data Modeling
- Value Chain Analysis
- Related Data Architecture

Data Quality Management
- Specification
- Analysis
- Measurement
- Improvement

Data Development
- Analysis
- Data Modeling
- Database Design
- Implementation

Meta Data Management
- Architecture
- Integration
- Control
- Delivery

Data Governance

Database Operations Management
- Acquisition
- Recovery
- Tuning
- Retention
- Purging

Document & Content Management
- Acquisition & Storage
- Backup & Recovery
- Content Mgmt.
- Retrieval
- Retention

Data Warehousing & Business Intelligence Management
- Architecture
- Implementation
- Training & Support
- Monitoring & Tuning

Reference & Master Data Management
- External Codes
- Internal Codes
- Customer Data
- Product Data
- Dimension Mgmt

Data Security Management
- Standards
- Classification
- Administration
- Authentication
- Auditing

Slide 16

DAMA DM BoK: Data Development

Definition: Designing, implementing, and maintaining solutions to meet the data needs of the enterprise.

Goals:
1. Identify and define data requirements.
2. Design data structures and other solutions to these requirements.
3. Implement and maintain solution components that meet these requirements.
4. Ensure solution conformance to data architecture and standards as appropriate.
5. Ensure the integrity, security, usability, and maintainability of structured data assets.

Inputs:
- Business Goals and Strategies
- Data Needs and Strategies
- Data Standards
- Data Architecture
- Process Architecture
- Application Architecture
- Technical Architecture

Suppliers:
- Data Stewards
- Subject Matter Experts
- IT Steering Committee
- Data Governance Council
- Data Architects and Analysts
- Software Developers
- Data Producers
- Information Consumers

Participants:
- Data Stewards and SMEs
- Data Architects and Analysts
- Database Administrators
- Data Model Administrators
- Software Developers
- Project Managers
- DM Executives and Other IT Management

Activities:
1. Data Modeling, Analysis and Solution Design (D)
   1. Analyze Information Requirements
   2. Develop and Maintain Conceptual Data Models
   3. Develop and Maintain Logical Data Models
   4. Develop and Maintain Physical Data Models
2. Detailed Data Design (D)
   1. Design Physical Databases
   2. Design Information Products
   3. Design Data Access Services
   4. Design Data Integration Services
3. Data Model and Design Quality Management
   1. Develop Data Modeling and Design Standards (P)
   2. Review Data Model and Database Design Quality (C)
   3. Manage Data Model Versioning and Integration (C)
4. Data Implementation (D)
   1. Implement Development / Test Database Changes
   2. Create and Maintain Test Data
   3. Migrate and Convert Data
   4. Build and Test Information Products
   5. Build and Test Data Access Services
   6. Validate Information Requirements
   7. Prepare for Data Deployment

Tools:
- Data Modeling Tools
- Database Management Systems
- Software Development Tools
- Testing Tools
- Data Profiling Tools
- Model Management Tools
- Configuration Management Tools
- Office Productivity Tools

Primary Deliverables:
- Data Requirements and Business Rules
- Conceptual Data Models
- Logical Data Models and Specifications
- Physical Data Models and Specifications
- Meta-data (Business and Technical)
- Data Modeling and DB Design Standards
- Data Model and DB Design Reviews
- Version Controlled Data Models
- Test Data
- Development and Test Databases
- Information Products
- Data Access Services
- Data Integration Services
- Migrated and Converted Data

Consumers:
- Data Producers
- Knowledge Workers
- Managers and Executives
- Customers
- Data Professionals
- Other IT Professionals

Activities: (P) – Planning (C) – Control (D) – Development (O) - Operational

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Data Modeling Fundamentals

- Data Management Contextual Overview
- Motivation
  - of systems/components
  - Data is not well understood
- Why data modeling & what is it?
  - Model represents our understanding of the
  - Fundamental, foundational system characteristics
  - Shared between system and human
- Fundamentals
  - The power of the purpose statement
  - Understanding data centric thinking
  - Data modeling compliments other architecture/ engineering techniques, as well as
  - Challenges beyond data modeling
- Take Aways, References & Q&A

Mark your calendar for our monthly #DataEd webinars

DATAVERSITY

Architecture: here, whether you like it or not

deviantart.com

- <u>All</u> organizations have architectures

  - Some are better **understood** and **documented** (and therefore more **useful** to the organization) than others
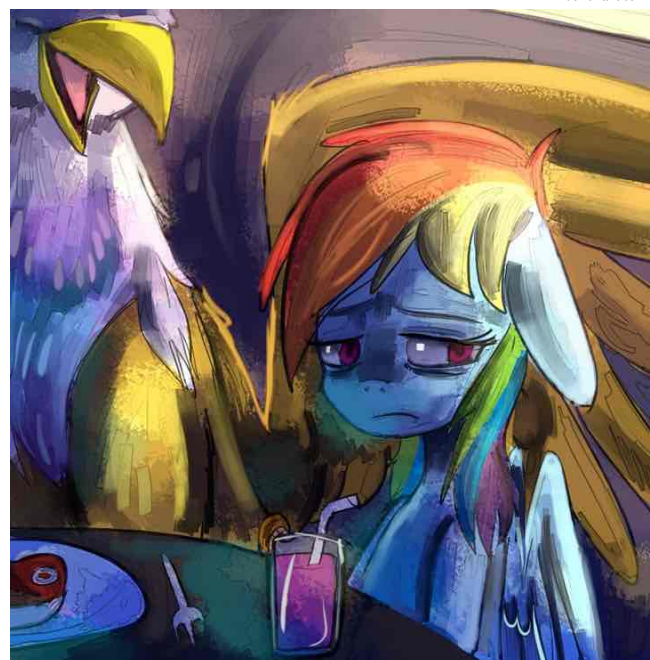
Data Architecture

- Architecture is higher level of abstraction
  - Understanding/integration focused

and

Models are also (literally) the translation between systems and people

Data Models

- Models more downward facing
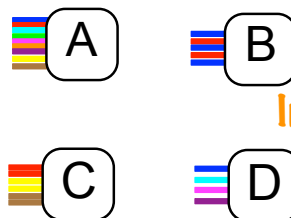 - Implementation/detail focused

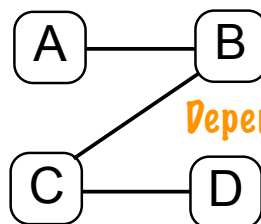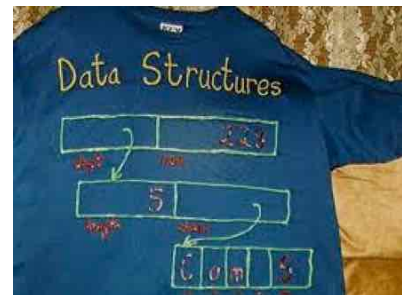http://www.architecturalcomponentsinc.com

---

How are components expressed as architectures?

- Details are organized into larger components

- Larger components are organized into models

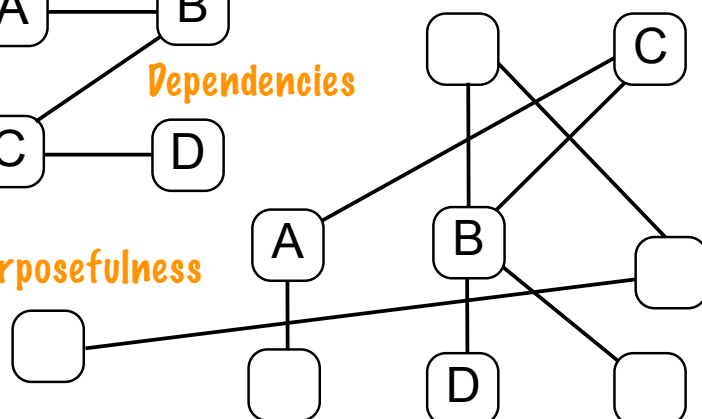- Models are organized into architectures (comprised of architectural components)

A B
C D
Intricate

Data Structures

A — B
C — D
Dependencies

Purposefulness

How are data structures expressed as architectures?

- **Attributes** are organized into entities/objects
  - Describe characteristics of "things" that someone cares to keep information about
  - Examples: color, size, sequence, media code, product descriptions

Intricate

- **Entities**/objects are organized into models
  - Combinations of attributes and entities are structured to represent information requirements
  - Entitles/objects are "things" whose information is managed in support of strategy

Dependencies

  - How the entitles interact
  - Relationships: accomplished by cooperating (sharing key information) Ex: An order is placed by one and only one customer
  - Poorly structured data, constrains organizational information delivery capabilities
  - Examples: persons, places, things

- **Models** are organized into architectures

Purposefulness

  - When building new systems, architectures are used to plan development
  - More often, data managers do not know what existing architectures are and - therefore - cannot make use of them in support of strategy implementation

---

Q: What is an Attribute?

club id #

- What does the existence of this attribute tell us?

  - Clubs need to be identified (#) separately from one another

  - Club-specific information is likely maintained

  - Some concept (organization) exists above the 'club level'

  - ...

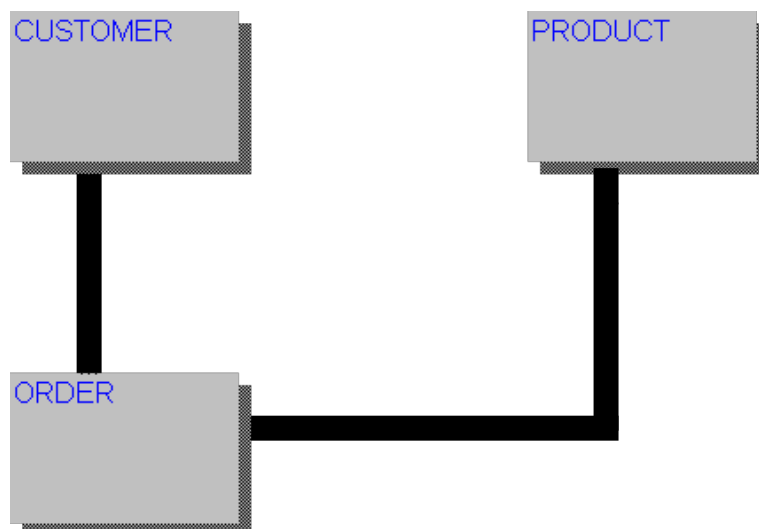# A: Attribute Definition

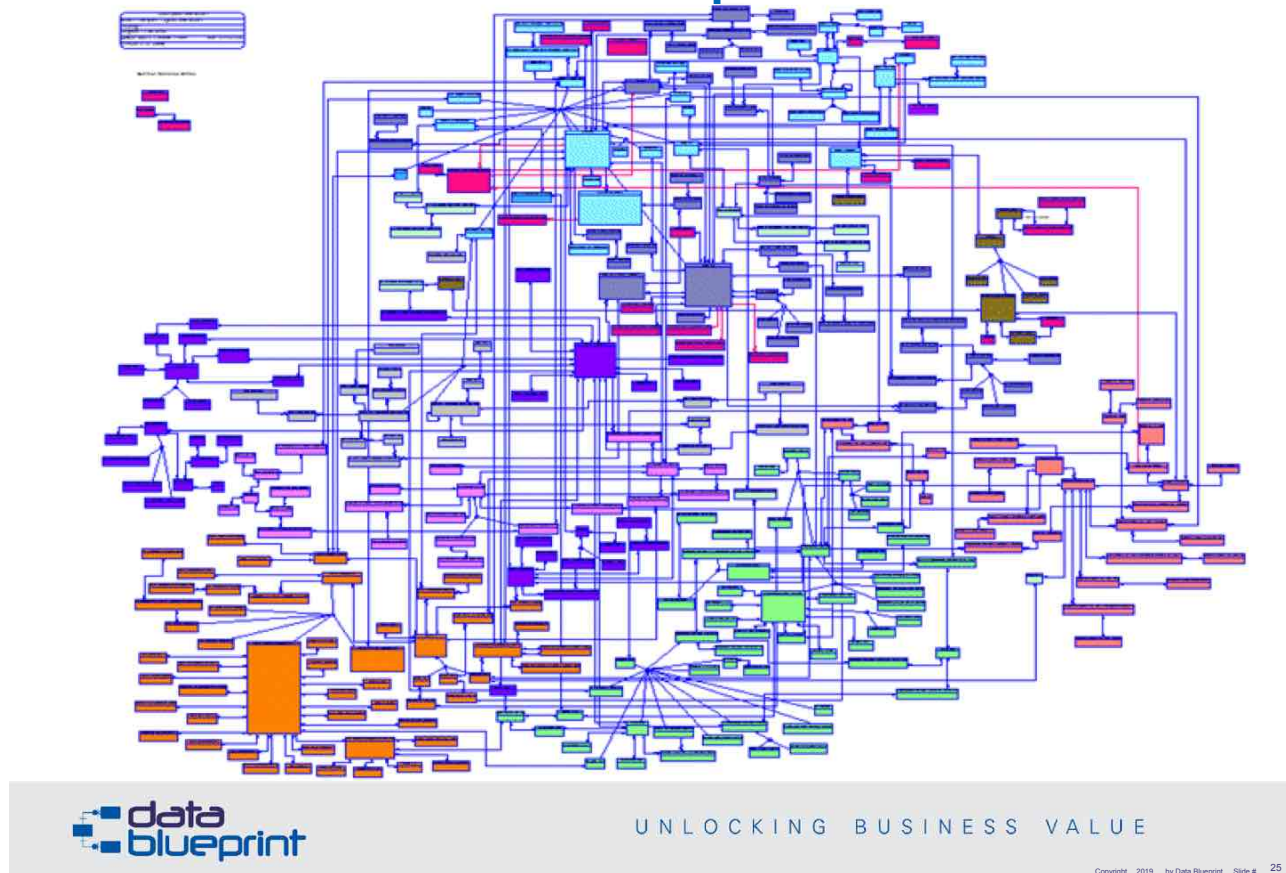- Attributes describe an entity and attribute values describe "instances of business things"

CLUB

club id #
current promotion
maximum period of obligation
number canceled ytd
number of members
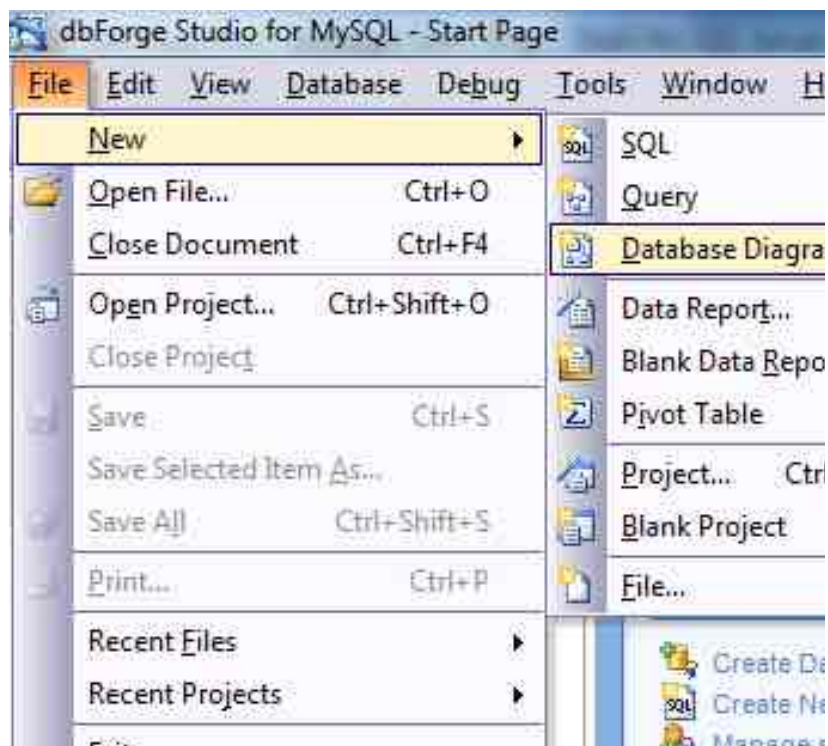total units sold for club

Entities organized into a model

CUSTOMER

PRODUCT

ORDER

Data architectures are comprised of data models

What do we teach IT professionals about data?

- 1 course

    – How to build a new database

- What impressions do IT professionals get from this education?

    – Data is a technical skill that is needed when developing new databases

What do we teach knowledge workers about data?

(NOTHING)

What percentage of the deal with it daily?
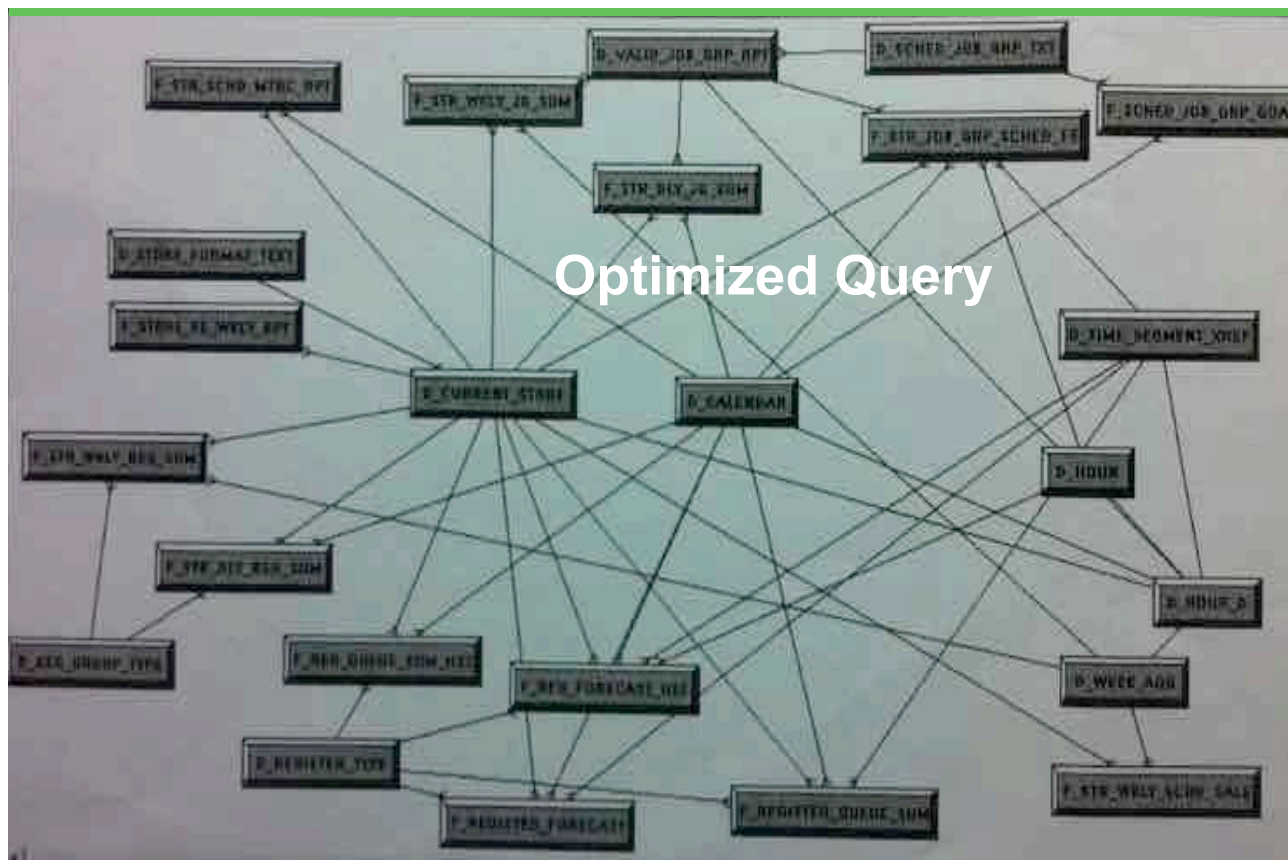
100%

Data Footprints

- SQL Server
  - 47,000,000,000,000 bytes
  - Largest table 34 billion records 3.5 TBs
- Informix
  - 1,800,000,000 queries/day
  - 65,000,000 tables / 517,000 databases
- Teradata
  - 117 billion records
  - 23 TBs for one table
- DB2
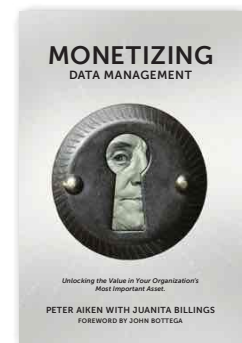 - 29,838,518,078 daily queries

Running Query

Optimized Query

Repeat 100s, thousands, millions of times ...

DEATH BY A THOUSAND CUTS
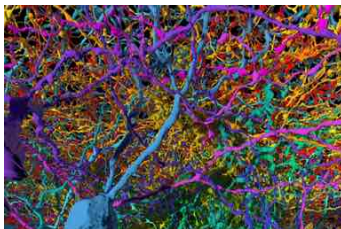
Lack of data coherence is a hidden expense

- How does maltreated data cost money?
- Consider the opposite question:
  - Were your systems explicitly designed to be integrated or otherwise work together?
  - If not then what is the likelihood that they will work well together?
- Organizations spend 20-40% of their IT budget evolving data - including:
  - Data migration
    - Changing the location from one place to another
  - Data conversion
    - Changing data into another form, state, or product
  - Data improving
    - Inspecting and manipulating, or re-keying data to prepare it for subsequent use   - John Zachman

COHERENCE

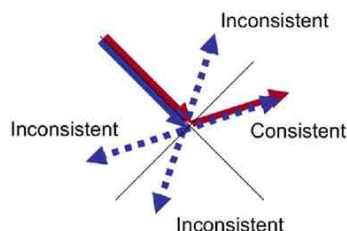MONETIZING
DATA MANAGEMENT

Unlocking the Value in Your Organization's Most Important Asset.

PETER AIKEN WITH JUANITA BILLINGS
FOREWORD BY JOHN BOTTEGA

As a topic, Data is ...

Complex & detailed

- Outsiders do not want to hear about or discuss any aspects of challenges/solutions
- Most are unqualified re: architecture/ engineering

Taught inconsistently

- Focus is on technology
- Business impact is not addressed

Not well understood

- (Re)learned by every workgroup
- Lack of standards/ poor literacy/ unknown dependencies

Wally Easton Playing Piano
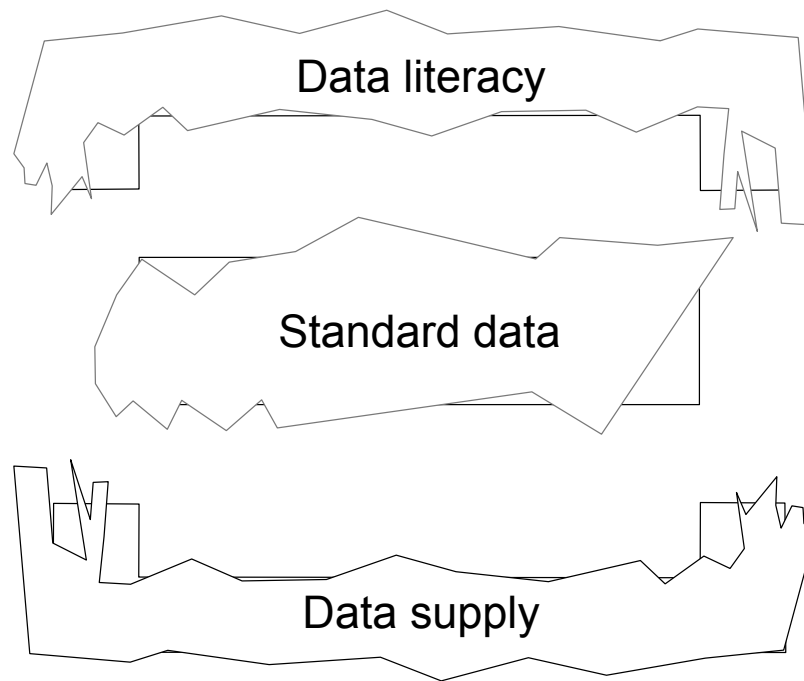https://www.youtube.com/watch?v=NNbPxSvII-Q

Making a Better Data Sandwich

Making a Better Data Sandwich

Data literacy

Standard data

Data supply

Making a Better Data Sandwich

Data literacy

Standard data

Data supply
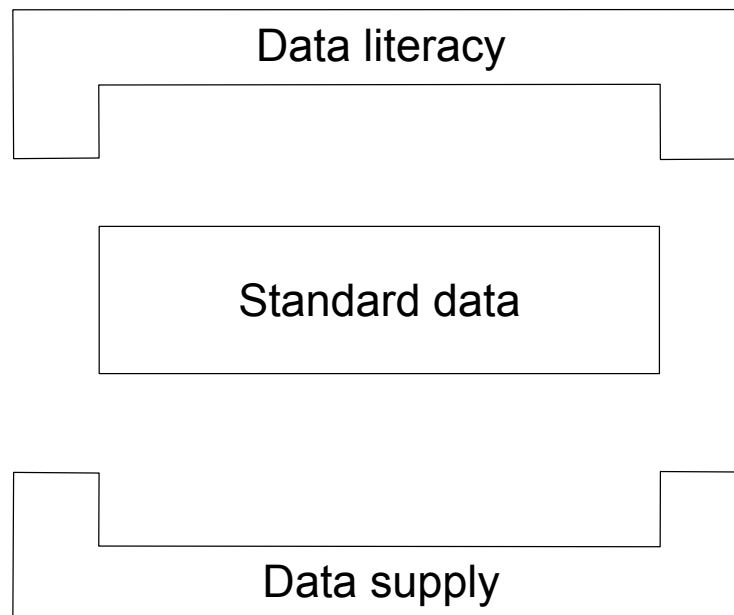
Making a Better Data Sandwich

This cannot happen without engineering and architecture!

Data literacy
Standard data
Data supply
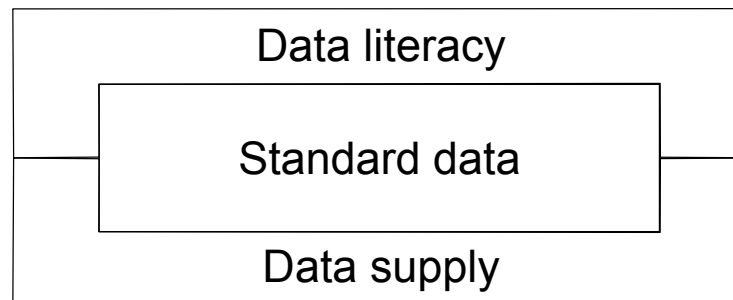
Quality engineering/ architecture work products do not happen accidentally!
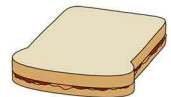
Making a Better Data Sandwich

This cannot happen without data engineering and architecture!

Data literacy
Standard data
Data supply
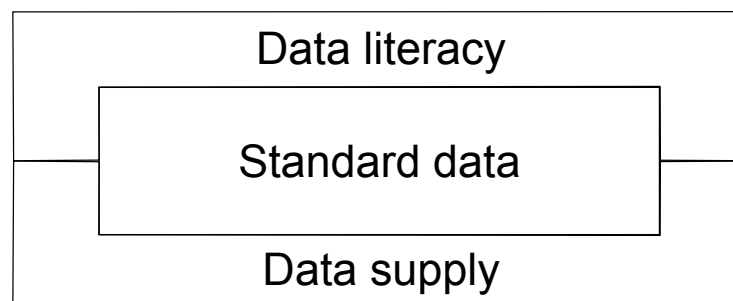
Quality data engineering/ architecture work products do not happen accidentally!

What is **this**?

- It is tall
- It has a clutch
- It was built in 1942
- It is cemented to the floor
- It is still in regular use!

Live the Adventure,
Honor the Legend

You cannot architect after implementation!

**Good Engineering/
Architectural
Foundation?**

Poor Foundation =

Unsuitable for Further Investment

Bad Data Decisions Spiral

```
┌─────────────────────┐      ┌─────────────────────┐
│  Business decision  │      │  Technical deci-    │
│  makers are not     │      │  sion makers are not│
│  data knowledgable  │      │  data knowledgable  │
└─────────────────────┘      └─────────────────────┘
              │                    │
              └────────┬───────────┘
                       ▼
        ┌──────────────────────────────┐
        │      Bad data decisions       │
        └──────────────────────────────┘
              │                    │
      ┌───────┘                    └────────┐
      ▼                                      ▼
┌─────────────────────┐      ┌─────────────────────┐
│  Poor treatment of  │      │      Poor           │
│  organizational data│      │      quality        │
│      assets         │      │      data           │
└─────────────────────┘      └─────────────────────┘
      │                                      │
      └──────────────┬───────────────────────┘
                     ▼
        ┌──────────────────────────────┐
        │  Poor organizational outcomes │
        └──────────────────────────────┘
```
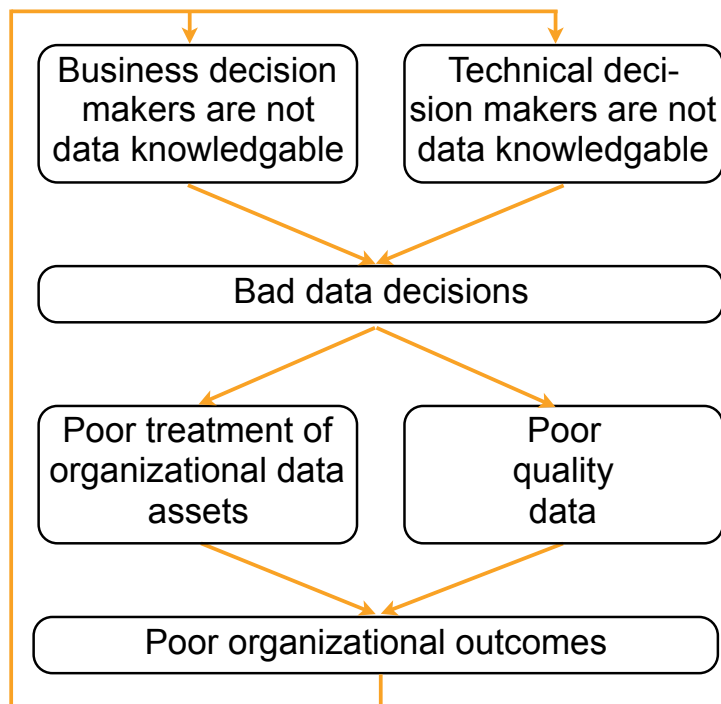
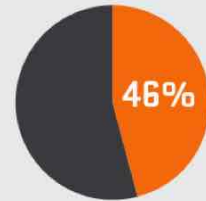**NEWS FLASH!**
46% of companies report they made an inaccurate business decision based on bad or outdated data. Bad data leads to bad business decisions. Companies need to be careful that their data is sound – especially when dealing with investors.

Like Comment Share

46%

Data Modeling Fundamentals

- Data Management Contextual Overview
- Motivation
  - of systems/components
  - Data is not well understood
- Why data modeling & what is it?
  - Model represents our understanding of the
  - Fundamental, foundational system characteristics
  - Shared between system and human
- Fundamentals
  - The power of the purpose statement
  - Understanding data centric thinking
  - Data modeling compliments other architecture/ engineering techniques, as well as
  - Challenges beyond data modeling
- Take Aways, References & Q&A

Mark your calendar for our monthly #DataEd webinars

DATAVERSITY

Data Modeling Definition

- **Modeling** = Analysis and design method used to
  - *Define and analyze data requirements*
  - *Design data structures that support these requirements*
- **Model** = set of data specifications and related diagrams that reflect requirements and designs
  - *Representation of something in our environment*
  - *Employs standardized text/symbols to represent data attributes (grouped into data elements) and the relationships among them*
  - *Integrated collection of specifications and related diagrams that represent data requirements and design*
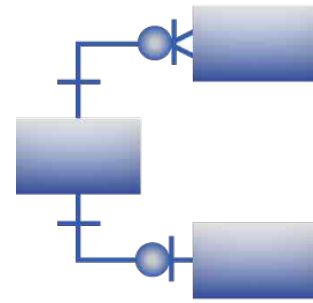
from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

---

Data Modeling

- **Modeling** = complex process involving interaction between people and with technology that don't compromise the integrity or security of the data

  - Good data models accurately express and effectively communicate data requirements and quality solution design

- **Modeling approach** (guided by 2 formulas):

  - Purpose + audience = deliverables

  - Deliverables + resources + time = approach

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Data Models Facilitate

- Formalization
  - Data model documents a single, precise definition of data requirements and data-related business rules

- Communication
  - Data model is a bridge to understanding data between people with different levels and types of experience.
  - Helps understand business area, existing application, or impact of modifying an existing structure
  - May also facilitate training new business and/or technical staff

- Scope
  - Data model can help explain the data concept and scope of purchased application packages
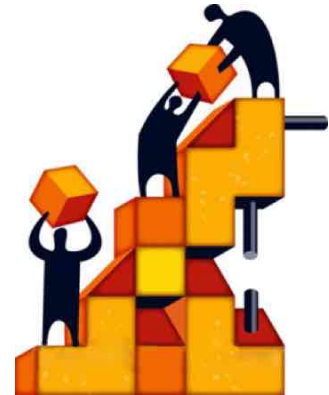
---

ANSI-SPARK 3-Layer Schema

1. **Conceptual** - Allows independent customized user views:
   - Each should be able to access the same data, but have a different customized view of the data.
2. **Logical** - This hides the physical storage details from users:
   - Users should not have to deal with physical database storage details. They should be allowed to work with the data itself, without concern for how it is physically stored.
3. **Physical** - The database administrator should be able to change the database storage structures without affecting the users' views:
   - Changes to the structure of an organization's data will be required. The internal structure of the database should be unaffected by changes to the physical aspects of the storage.

ANSI-SPARC Architecture for Databases

Users

external level (View) — multiple user's views

conceptual level (Schema) — Community view of DB

Physical representation

internal level (Schema)

Database (Physical level)
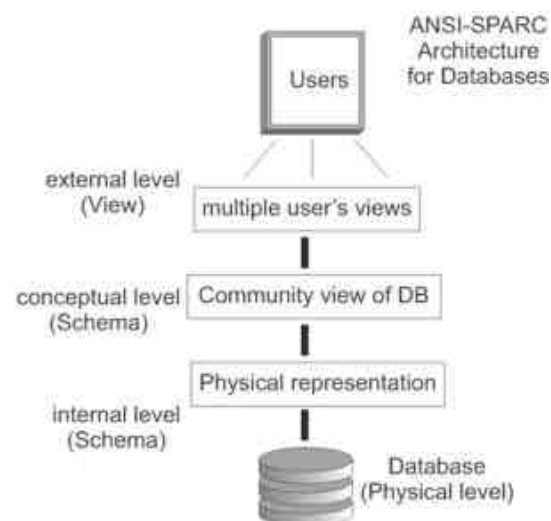
For example, a changeover to a new DBMS technology. The database administrator should be able to change the conceptual or global structure of the database without affecting the users.

Families of Modeling Notation Variants

Chen Style

Ordinality - describes the minimum (optional vs mandatory) → M:N ← Cardinality - describes the maximum
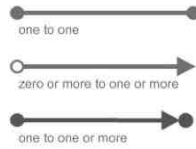
1:N (n=0,1,2,3...) one to zero or more

M:N (m and n=0;1,2,3...) zero or more to zero or more (many to many)

1:1 one to one

Company — 1 — ◇ — N — Employee — ◇ M:N — Projects

Bachman Style

●——————● one to one

○———————▶ zero or more to one or more

●———————▶ one to one or more

Company → ◇ → Employee → ◇ → Projects

Martin Style

1 - one, and only one (mandatory)

* - many (zero or more - optional)
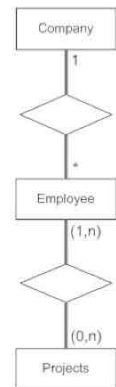
1...* - one or more (mandatory)

0...1 - zero or one (optional)

(0,1) - zero or one (optional)

(1,n) - one or more (mandatory)

(0,n) - zero or more (optional)

(1,1) - one and only one (mandatory)

Company — 1 — ◇ — * — Employee — (1,n) — ◇ — (0,n) — Projects

Information Engineering

———○≺	Eventually One, More
———○|	Eventually One
———|—	Exactly One
———○≺	Zero, or More
———≺	One or More
———○|	Zero or One

✔
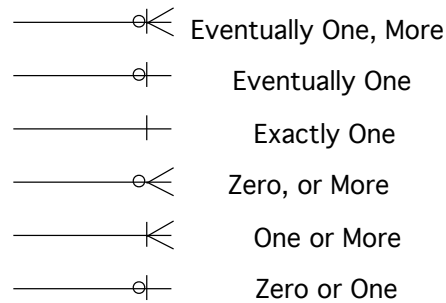
Just Pick ONE!

What is a Relationship?

• Natural associations between two or more entities

Ordinality & Cardinality

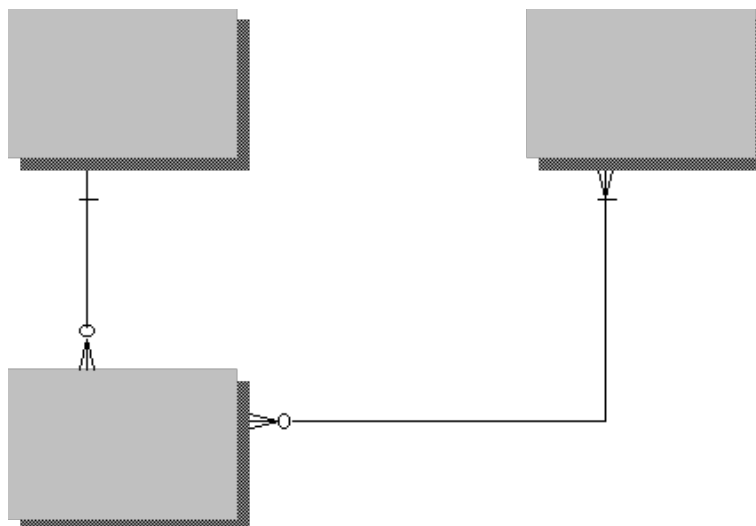- Defines mandatory/optional relationships using minimum/ maximum occurrences from one entity to another

CUSTOMER

PRODUCT

An order is placed by one and only one customer

A customer places zero or more orders

An order contains at least one or more products

ORDER

A product is contained on zero or more orders
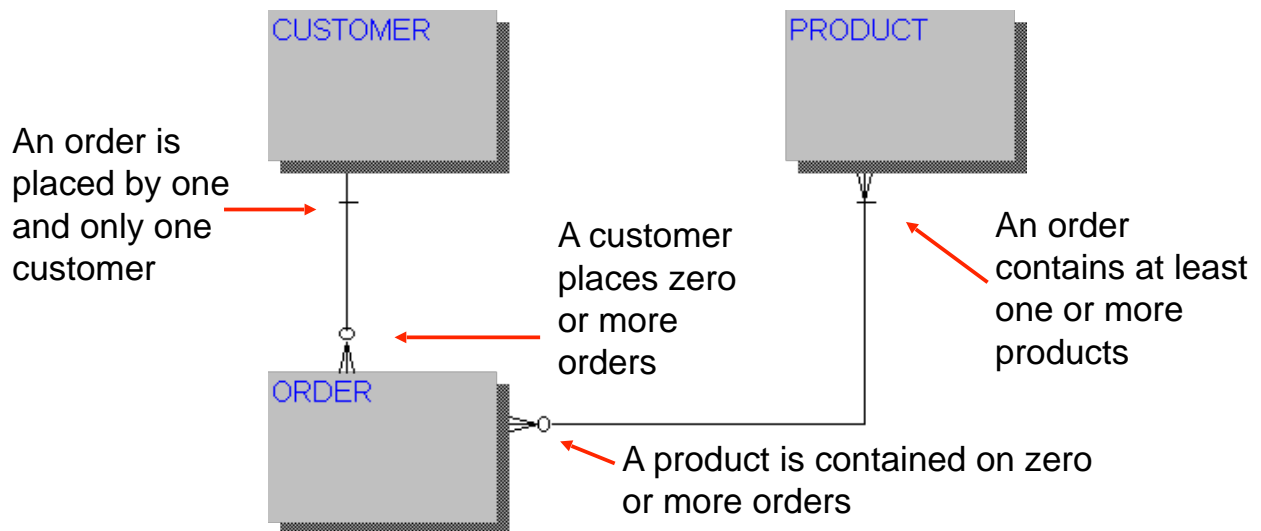
Q: What is the proper relationship for these entities?

MEMBER

CLUB

A: a relationship for these entities
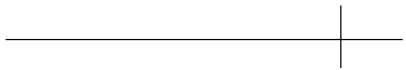
—————————|— Exactly One (mandatory)

—————————< One or Many (mandatory)
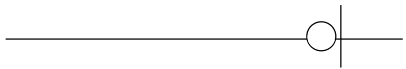
—————————○|— Eventually One (optional)
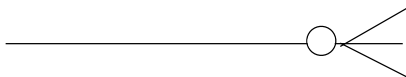
—————————○< Zero, or Many (optional)

—————————○< Eventually One or Many (optional)

Rigid Data Structure

```
Person                              Job Class
  |                                    |  |
  |        BR1) One EMPLOYEE           |  |
  |        can be associated with one  |  |
  |              PERSON                |  |
Manual                                 |  |
Moon Lighting                          |  |
  |        BR2) One EMPLOYEE can be     |  |
  |        associated with one POSITION |  |
  |                                     |  |
  ○                                     |  |
Employee ○—— Manual Job Sharing ——|— Position
```
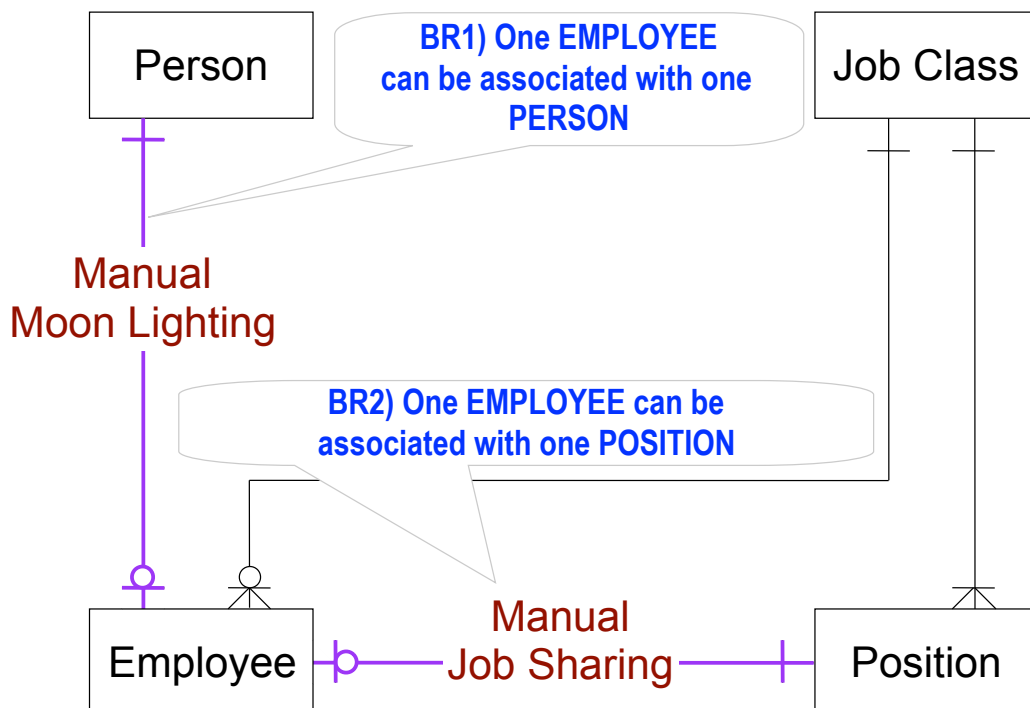
Flexible data structure
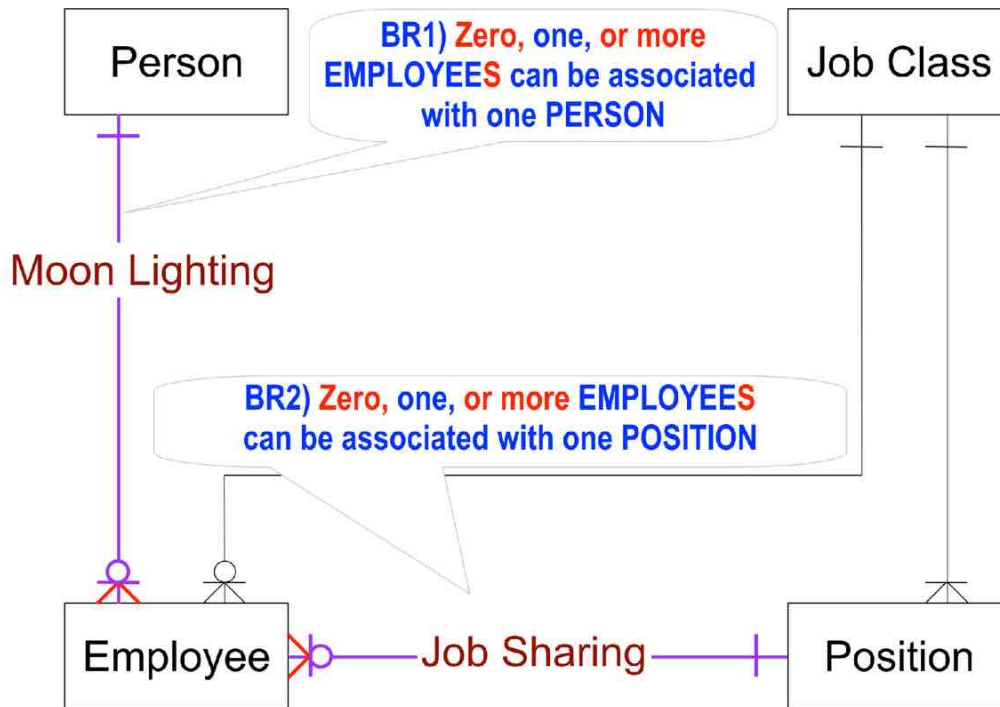
Everyone Shares Understanding

More flexible data structure

Less flexible data structure
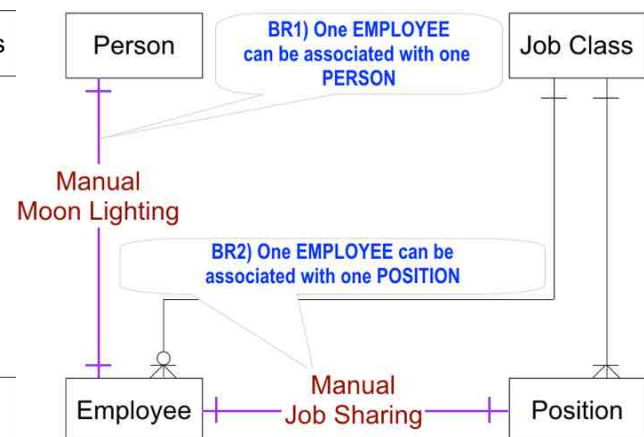
(Requires 2 structural loops more than the more flexible data structure)

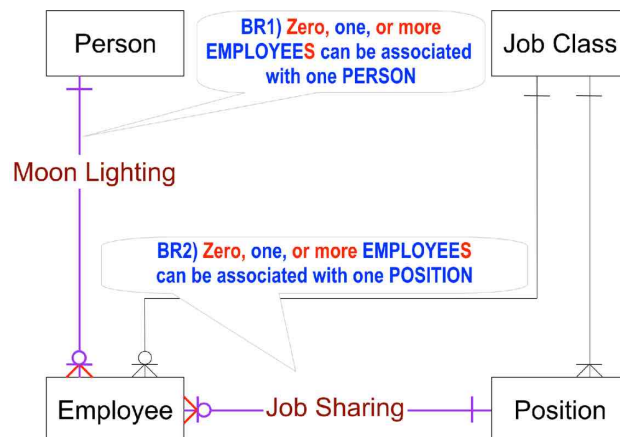## *Data structures must be specified prior software development/acquisition*

Understanding

- Definition:

    - 'Understanding an architecture'

    - Documented and articulated as a digital blueprint illustrating the commonalities and interconnections among the architectural components

    - Ideally the understanding is shared by systems and humans

Modeling Procedures

1. Identify entities

2. Identify key for each entity

3. Draw rough draft of entity relationship data model

4. Identify data attributes

5. Map data attributes to entities
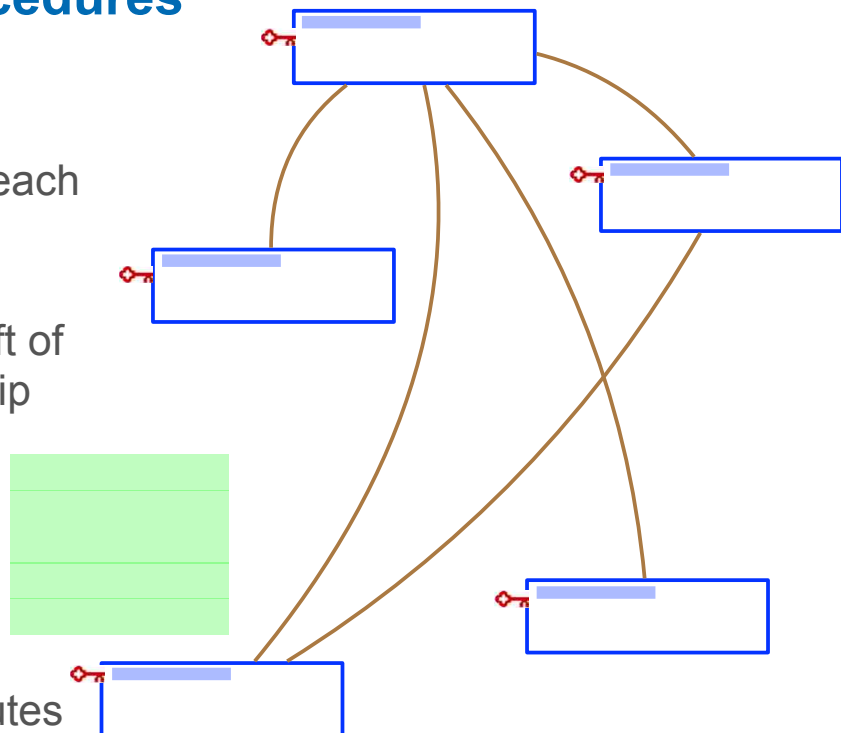
Models Evolution is good, at first ...
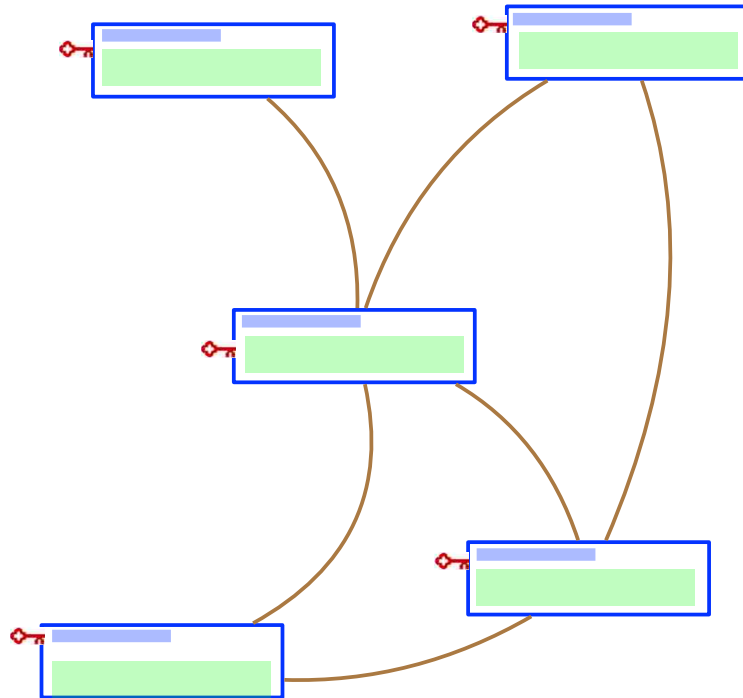
Relative use of time allocated to tasks during Modeling

Activity	Preliminary activities	Modeling cycles	Wrapup activities
Evidence collection & analysis	Collection		Analysis
Project coordination requirements	Declining coordination requirements		
Target system analysis		Increasing amounts of taget system analysis	
Modeling cycle focus	Refinement		Validation
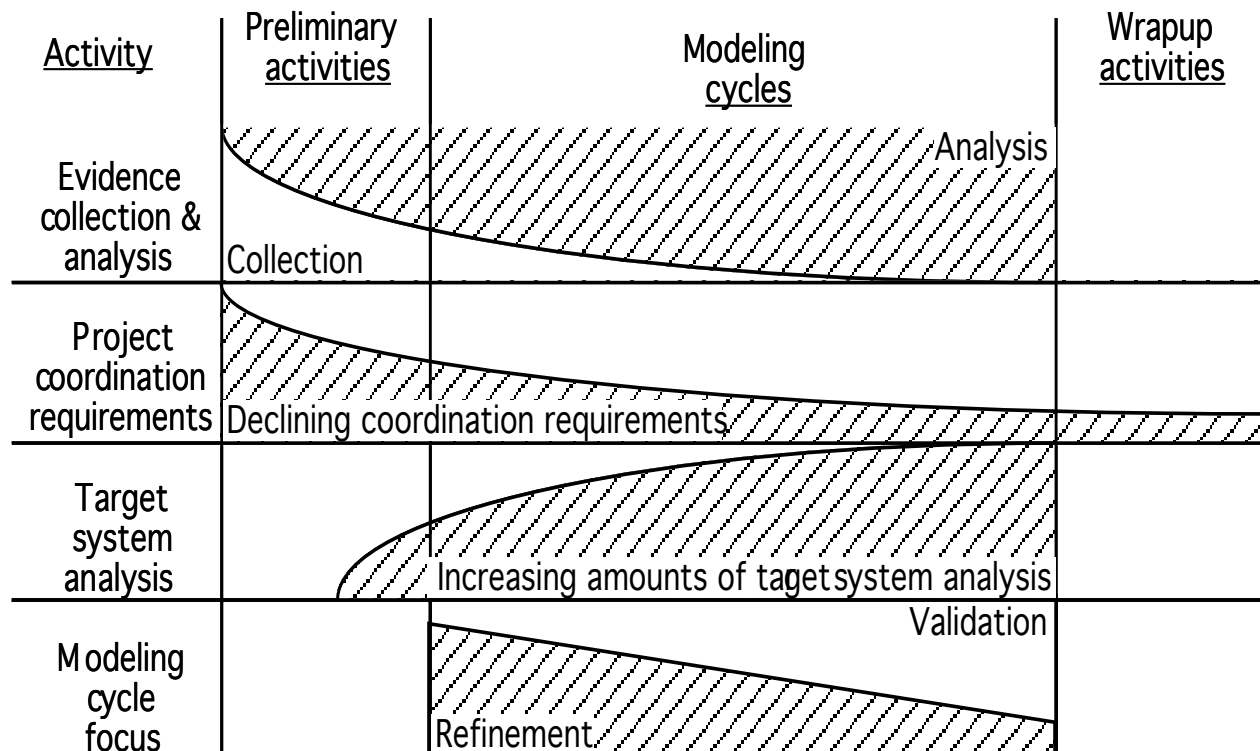
Don't Tell Them You Are Modeling!

- Just write some stuff down

- Then arrange it

- Then make some appropriate connections between your objects

Data Modeling Fundamentals

- Data Management Contextual Overview
- Motivation
  - of systems/components
  - Data is not well understood
- Why data modeling & what is it?
  - Model represents our understanding of the
  - Fundamental, foundational system characteristics
  - Shared between system and human
- Fundamentals
  - The power of the purpose statement
  - Understanding data centric thinking
  - Data modeling compliments other architecture/ engineering techniques, as well as
  - Challenges beyond data modeling
- Take Aways, References & Q&A

Mark your calendar for our monthly #DataEd webinars

DATAVERSITY

Each model has a purpose

Data Models are Developed in Response to Organizational Needs

Organizational Needs

become instantiated
and integrated into an

Data Models

authorizes and
articulates

Information System
Requirements

satisfy specific organizational needs
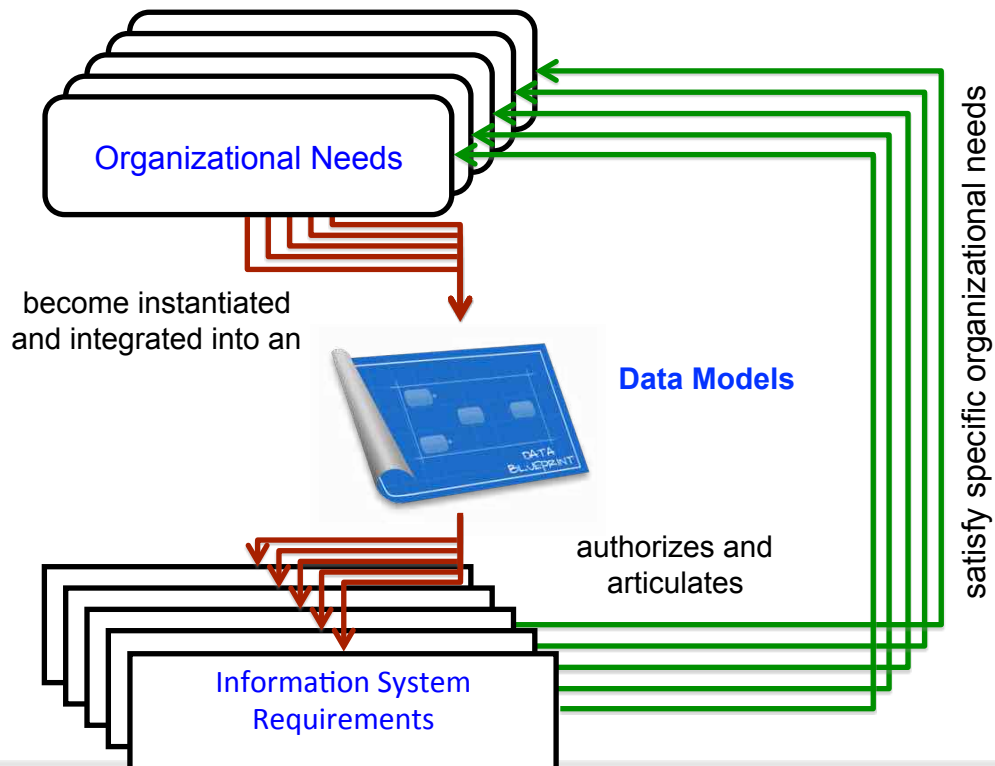
Standard definition reporting does not provide conceptual context

Bed

Something you sleep in
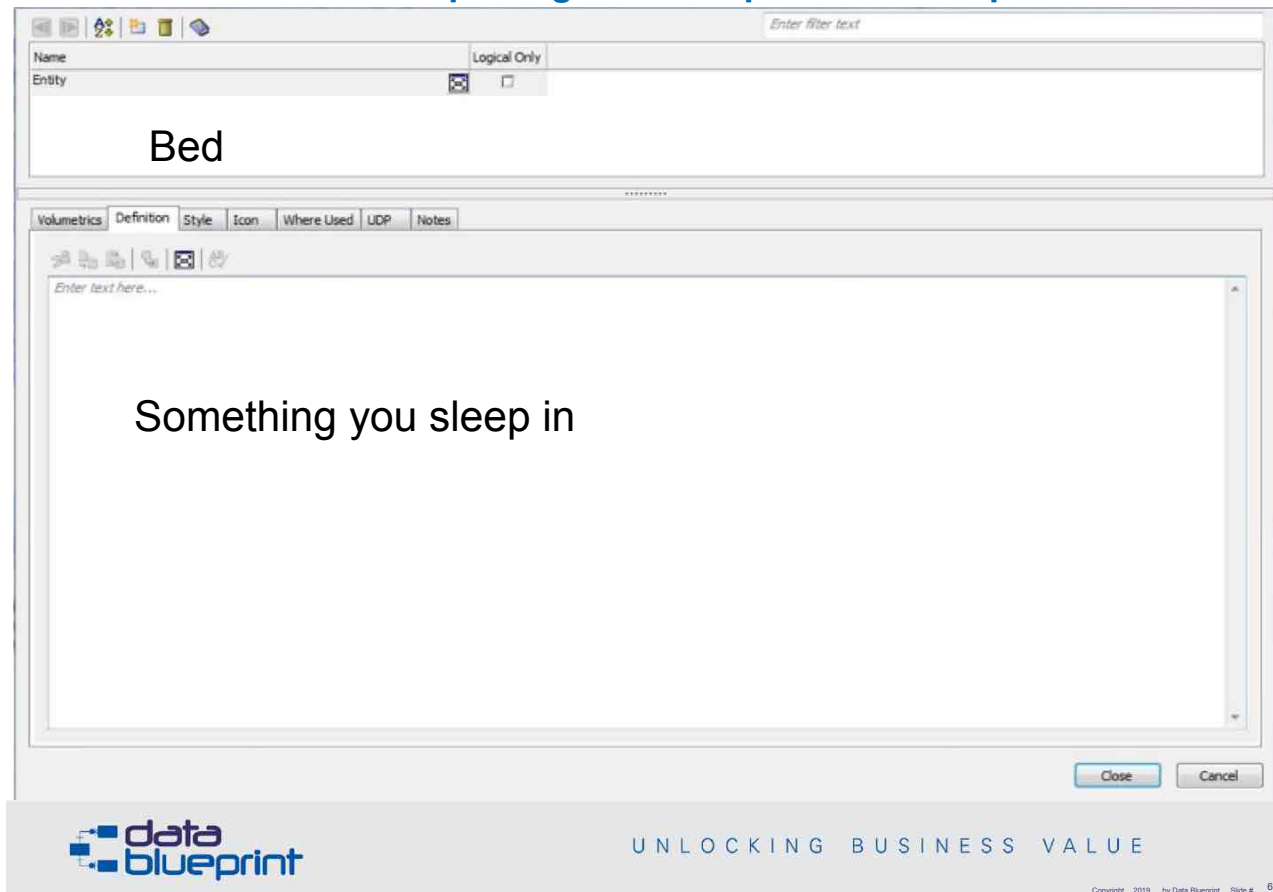
Keep them focused on data model purpose

- The reason we are locked in this room is to:
  – Mission: *Understand formal relationship between soda and customer*
    - Outcome: Walk out the door with a data model this relationship
  – Mission: *Understand the characteristics that differ between our hospital beds*
    - Outcome: We will walk out the door when we identify the top three traits that represent the brand.
  – Mission: *Could our systems handle the following business rule tomorrow?*
  – "Is job-sharing permitted?"
    - Outcomes: Confirm that it is possible to staff a position with multiple employees effective tomorrow

Bed
Entity: BED
Purpose: This is a substructure within the room substructure of the facility location. It contains information about beds within rooms.
Attributes: Bed.Description
              Bed.Status
              Bed.Sex.To.Be.Assigned
              Bed.Reserve.Reason
Associations: >0-+ Room
Status: Validated

How does our perspective change: *the primary means of tracking a patient*
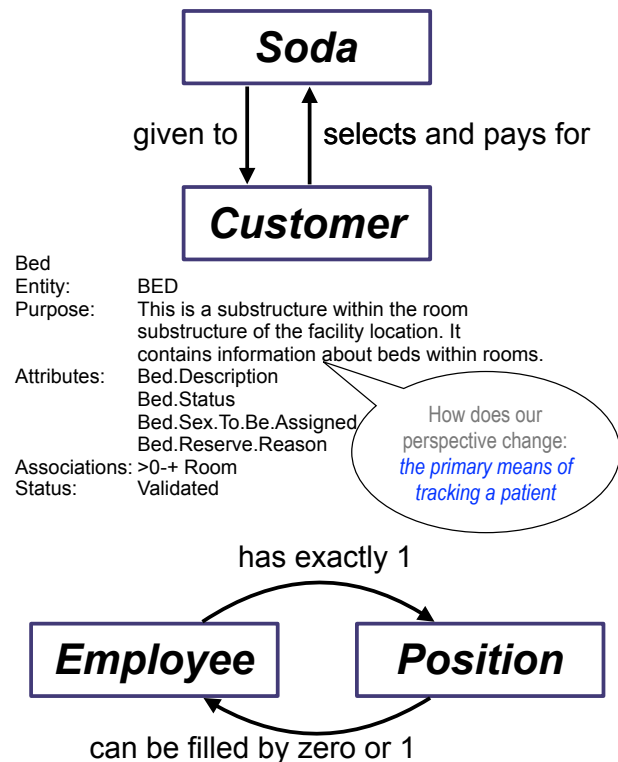
The Power of the Purpose Statement

Entity:	BED
Data Asset Type:	Principal Data Entity
Purpose:	This is a substructure within the <u>room</u> substructure of the <u>facility location</u>. It contains information about <u>beds</u> within <u>rooms</u>.
Source:	Maintenance Manual for File and Table Data (Software Version 3.0, Release 3.1)
Attributes:	<u>Bed.Description</u> <u>Bed.Status</u> <u>Bed.Sex.To.Be.Assigned</u> <u>Bed.Reserve.Reason</u>
Associations:	>0-+ <u>Room</u>
Status:	Validated

- A purpose statement describing why the organization is maintaining information about this business concept
- Sources of information about it
- A partial list of the attributes or characteristics of the entity
- Associations with other data items; this one is read as "One room contains zero or many beds"
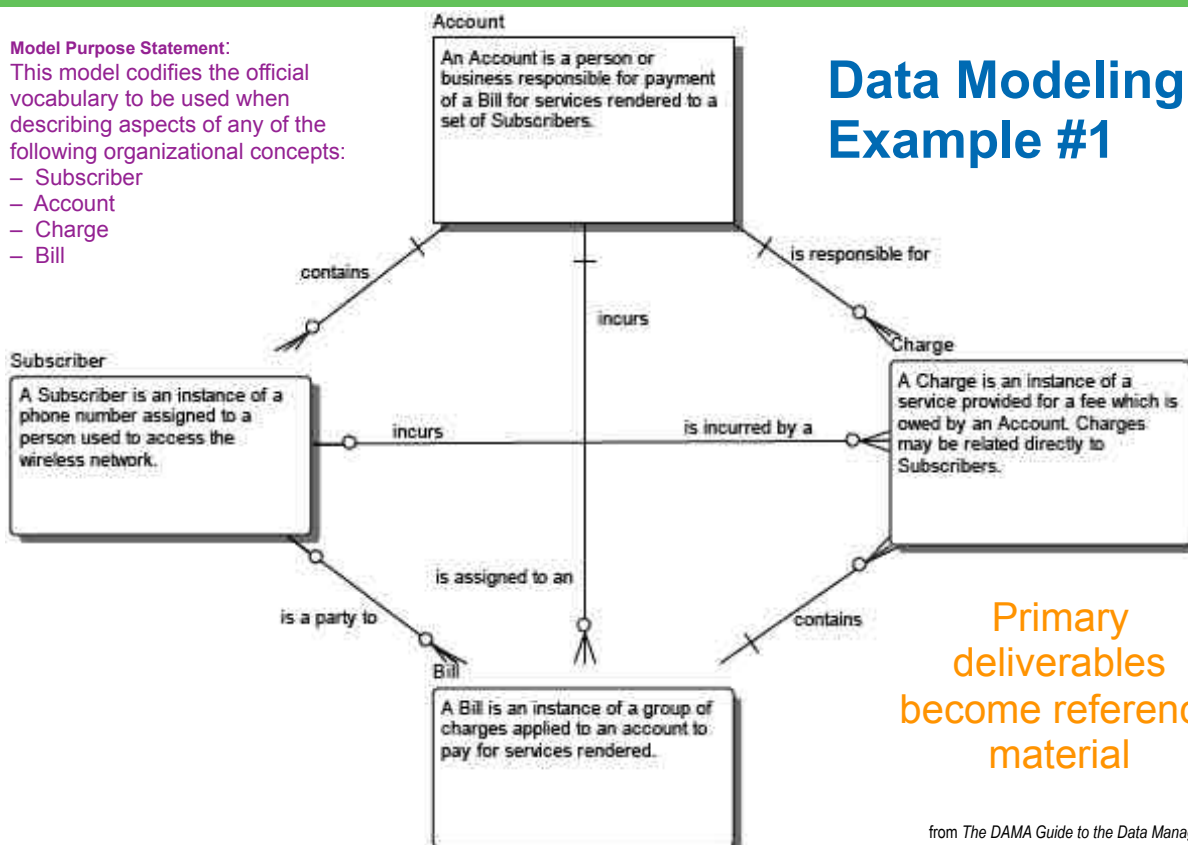

**Model Purpose Statement:**
This model codifies the official vocabulary to be used when describing aspects of any of the following organizational concepts:
– Subscriber
– Account
– Charge
– Bill

Data Modeling Example #1

Account
An Account is a person or business responsible for payment of a Bill for services rendered to a set of Subscribers.

Subscriber
A Subscriber is an instance of a phone number assigned to a person used to access the wireless network.

Charge
A Charge is an instance of a service provided for a fee which is owed by an Account. Charges may be related directly to Subscribers.

Bill
A Bill is an instance of a group of charges applied to an account to pay for services rendered.

Primary deliverables become reference material

from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Data Modeling Example #2

fuel

customer

rent-rate

rental agreement

auto

phone-call

phone-rate

phone-unit

repair history
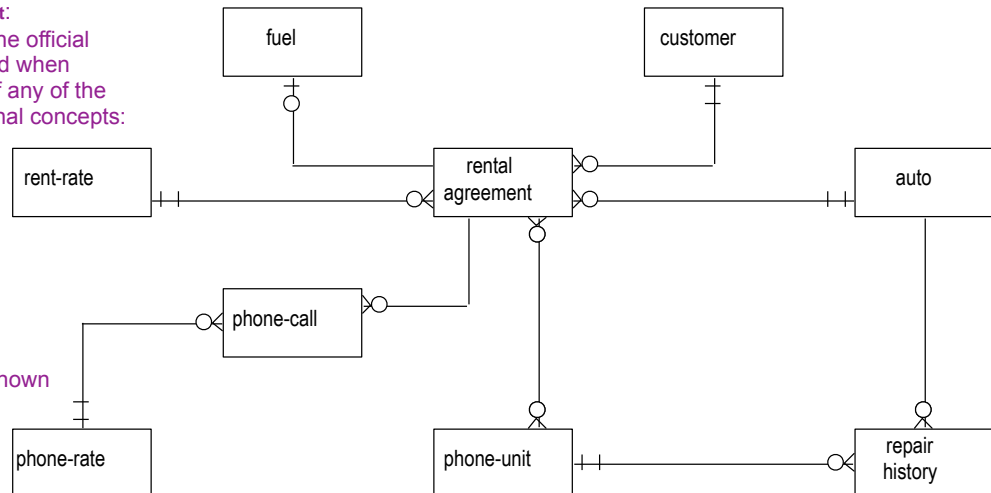
Interpretations:

1. Car rental company
2. Rental agreement is central
3. No direct connection between customer and contract
4. Contract must have a customer
5. Nothing structural prevents autos from being rented to multiple customers
6. Phone units are tied to rentals

Source: Chikofsky 1990

Data Modeling Example #3

- Sales commission-based pricing information
- Difficult to change a customer address
- Price not included in the catalog
- Easy to implement variable pricing - difficult to implement standard pricing - is standard pricing implemented
- Sales person information is not directly tied to the order
- Do sales people sell things that are shipped quickly so they get their commission quicker?
- Nothing prohibits a sales from having multiple sales persons
- Multiple invoices are allowed for a single order
- Partial shipment is allowed
- Data base cannot tell what part of an order the invoice pertains to

SALESPERSON

salesperson name	commission rate

INVOICE

invoice #	amount	date paid

ORDER

order #	date	customer #	customer name	address

LINE ITEM

item #	order #	quantity	price

CATALOG

item #	supplier	description	quantity on hand	cost

DISPOSITION Data Map

This model codifies the official
vocabulary to be used when
describing disposition related organizational concepts:
– user
– admission
– discharge
– encounter
– facility
– provider
– diagnosis
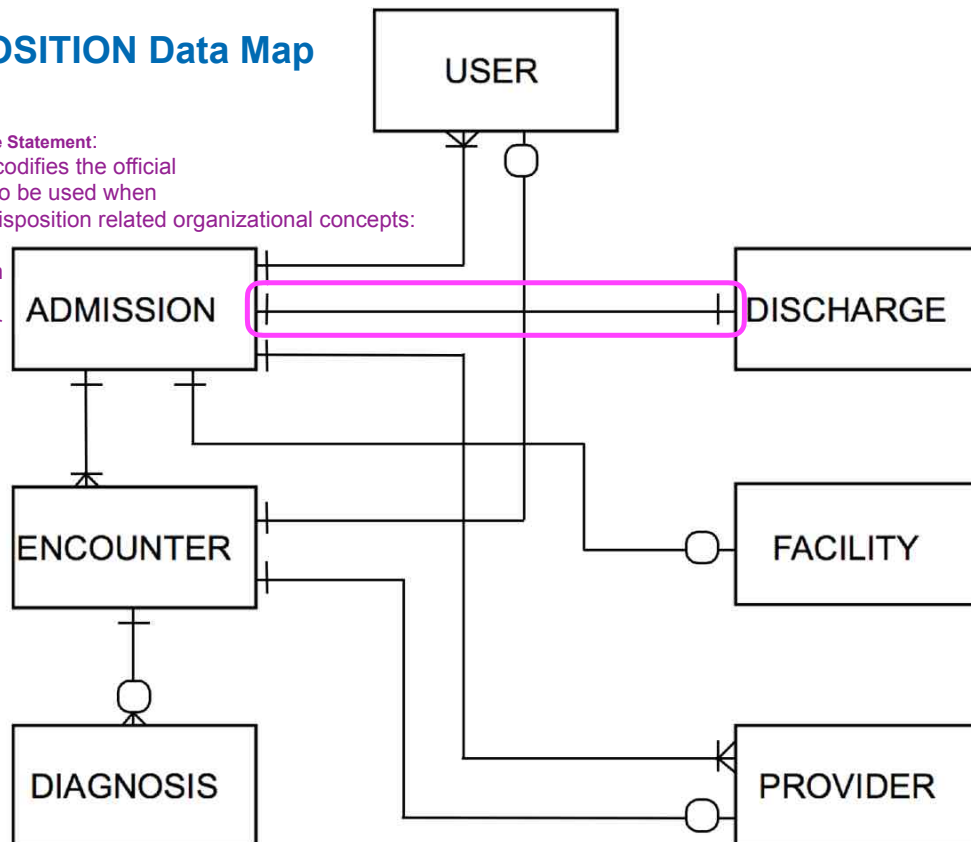
Data Model #4: DISPOSITION

- At least one but possibly more system USERS enter the DISPOSITION facts into the system.
- **An ADMISSION is associated with one and only one DISCHARGE.**
- An ADMISSION is associated with zero or more FACILITIES.
- An ADMISSION is associated with zero or more PROVIDERS.
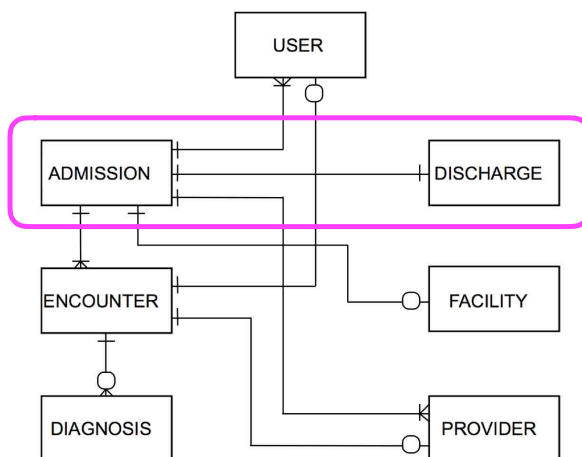
- An ADMISSION is associated with one or more ENCOUNTERS.
- An ENCOUNTER may be recorded by a system USER.
- An ENCOUNTER may be associated with a PROVIDER.
- An ENCOUNTER may be associated with one or more DIAGNOSES.

Death must be a disposition code!

ADMISSION	Contains information about patient admission history related to one or more inpatient episodes
DIAGNOSIS	Contains the International Disease Classification (IDC) of code representation and/or description of a patient's health related to an inpatient code
DISCHARGE	A table of codes describing disposition types available for an inpatient at a FACILITY
ENCOUNTER	Tracking information related to inpatient episodes
FACILITY	File containing a list of all facilities in regional health care system
PROVIDER	Full name of a member of the FACILITY team providing services to the patient
USER	Any user with access to create, read, update, and delete DISPOSITION data

IT Project or Application-Centric Development

- In support of strategy, organizations implement IT projects
- Data/information are typically considered within the scope of IT projects
- Problems with this approach:
  - Ensures data is formed to the applications and not around the organizational-wide information requirements
  - Process are narrowly formed around applications
  - Very little data reuse is possible
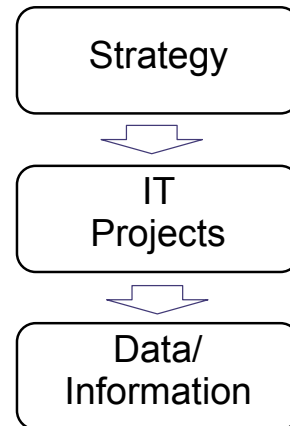
Strategy
↓
IT Projects
↓
Data/ Information

Data-Centric Development

- In support of strategy, the organization develops specific, shared data-based goals/objectives
- These organizational data goals/ objectives drive the development of specific IT projects with an eye to organization-wide usage
- Advantages of this approach:
  - Data/information assets are developed from an organization-wide perspective
  - Systems support organizational data needs and compliment organizational process flows
  - Maximum data/information reuse

Strategy
↓
Data/ Information
↓
IT Projects

theDataDoctrine.com

We are uncovering better ways of developing
IT systems by doing it and helping others do it.
Through this work we have come to value:

Data programmes preceding software development
Stable data structures preceding stable code
Shared data preceding completed software
Data reuse preceding reusable code

---

theDataDoctrine.com

We are uncovering better ways of developing
IT systems by doing it and helping others do it.
Through this work we have come to value:

Data programmes preceding software development
Stable data structures preceding stable code
Shared data preceding completed software
Data reuse preceding reusable code

That is, while there is value in the items on
the right, we value the items on the left more.

Typically Managed Architectures
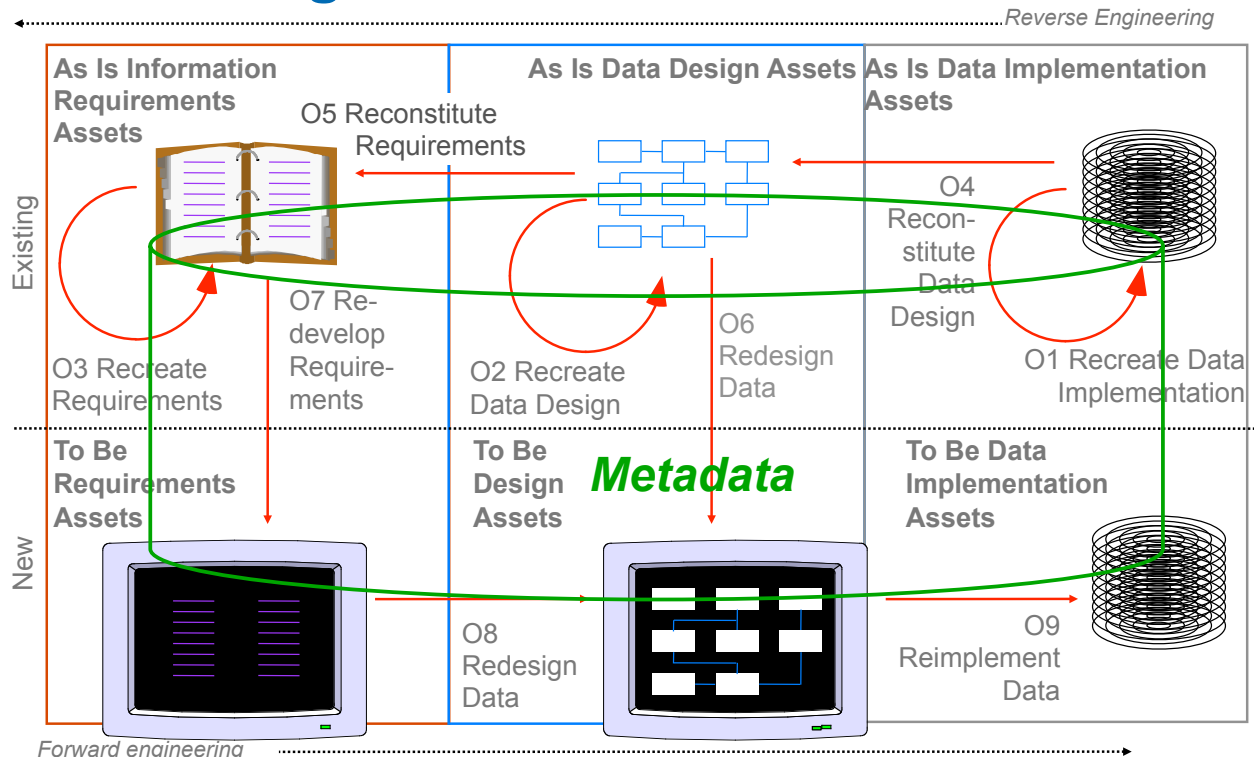
Technical committees

- Process Architecture
  - Arrangement of inputs -> transformations = value -> outputs
  - Typical elements: Functions, activities, workflow, events, cycles, products, procedures
- Systems Architecture
  - Applications, software components, interfaces, projects
- Business Architecture
  - Goals, strategies, roles, organizational structure, location(s)
- Security Architecture
  - Arrangement of security controls relation to IT Architecture
- Technical Architecture/Tarchitecture
  - Relation of software capabilities/technology stack
  - Structure of the technology infrastructure of an enterprise, solution or system
  - Typical elements: Networks, hardware, software platforms, standards/protocols
- Data/Information Architecture
  - Arrangement of data assets supporting organizational strategy
  - Typical elements: specifications expressed as entities, relationships, attributes, definitions, values, vocabularies

Modeling in Various Contexts

Reverse Engineering

Existing

As Is Information Requirements Assets

O5 Reconstitute Requirements

As Is Data Design Assets

As Is Data Implementation Assets

O4 Reconstitute Data Design

O7 Redevelop Requirements

O3 Recreate Requirements

O2 Recreate Data Design

O6 Redesign Data

O1 Recreate Data Implementation

New

To Be Requirements Assets

To Be Design Assets

Metadata

To Be Data Implementation Assets

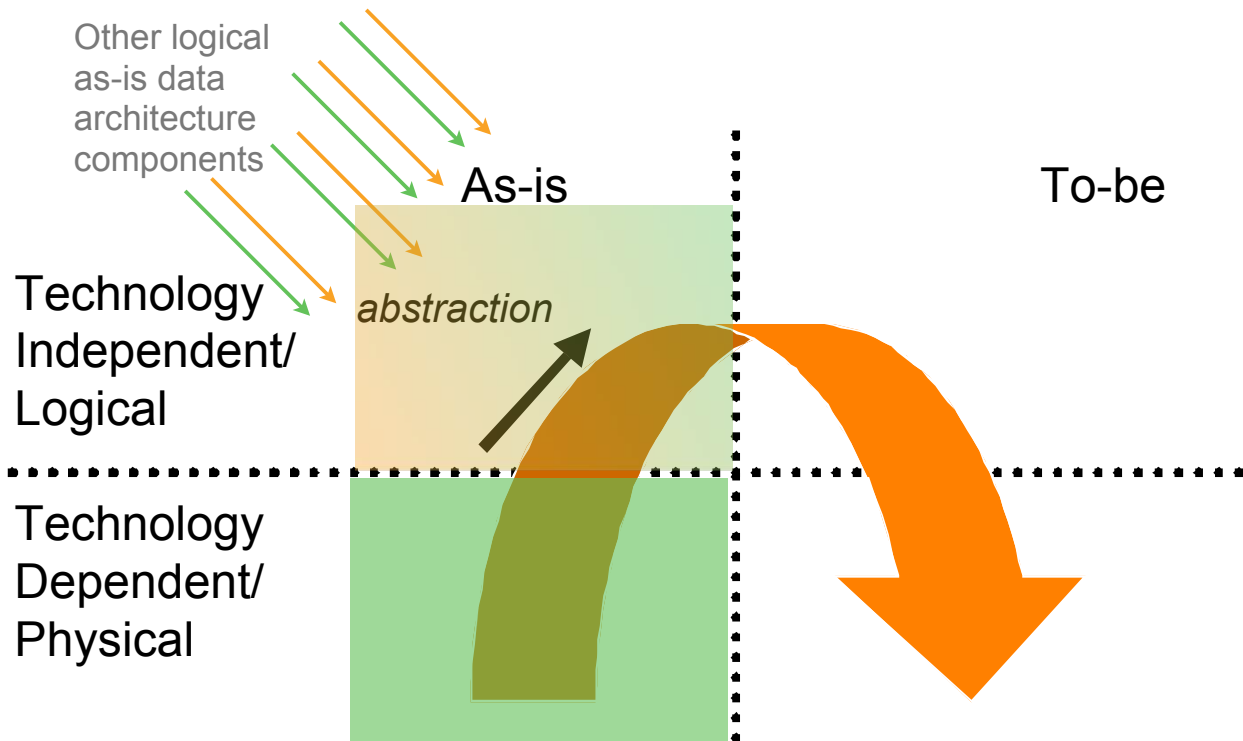O8 Redesign Data

O9 Reimplement Data

Forward engineering

Information Architecture Component Reengineering Options

O-1   data implementation (e.g., by recreating descriptions of implemented file layouts);

O-2   data designs (e.g., by recreating the logical system design layouts); or

O-3   information requirements (e.g., by recreating existing system specifications and business rules).

O-4   data design assets by examining the existing data implementation (when appropriate O-1 can facilitate O-4); and

O-5   system information requirements by reverse engineering the data design O-4. (Note: if the data design doesn't exist O-4 must precede O-5.)

O-6   transforming as is data design assets, yielding improved to be data designs that are based on reconstituted data design assets produced by O-2 or O-4 and (possibly O-1);

O-7   transforming as is system requirements into to be system requirements that are based on reconstituted system requirements produced by O-3 or O-5 and (possibly O-2);

O-8   redesigning to be data design assets using the to be system requirements based on reconstituted system requirements produced by O-7; and

O-9   re-implementing system data based on data redesigns produced by O-6 or O-8.

Model Evolution Framework

Every change can be mapped to a transformation in this framework!

Goal

To Be

As Is

Validated

Not Validated

Conceptual   Logical   Physical
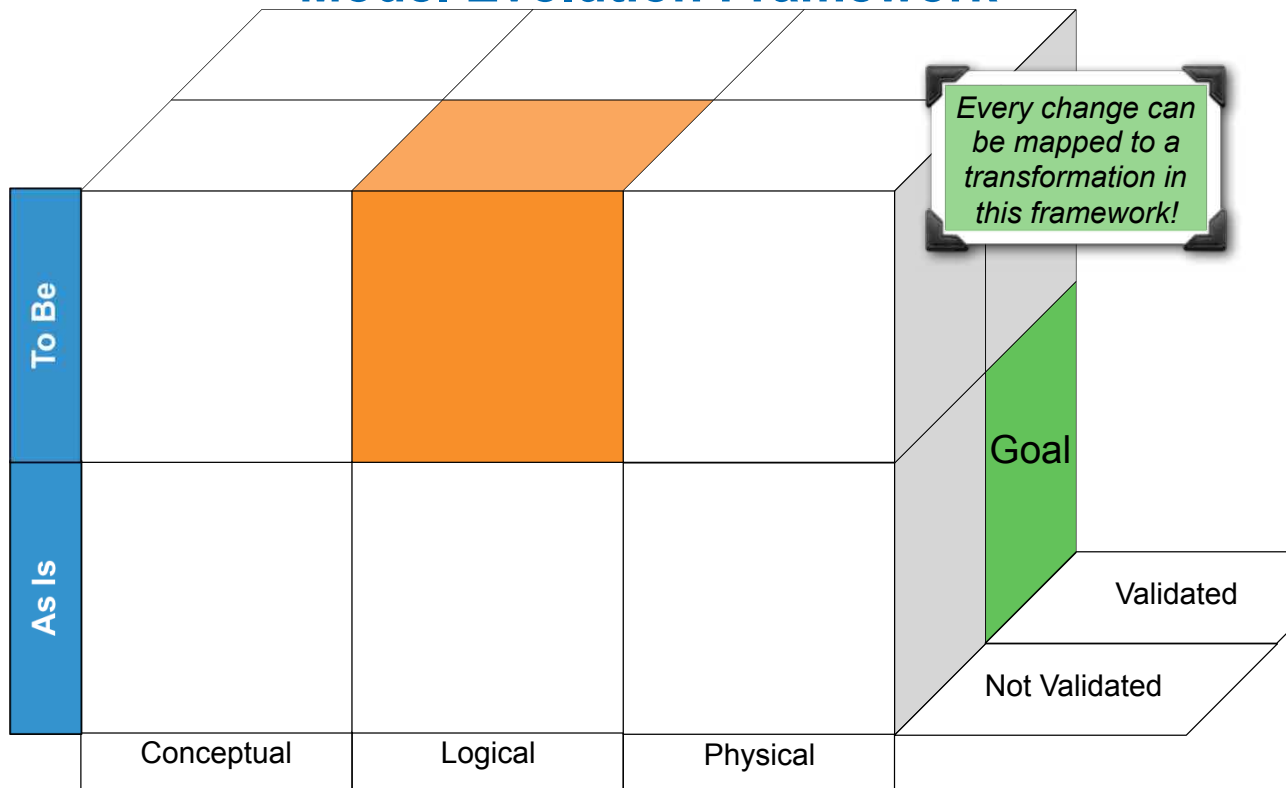
Model Evolution (better explanation)

Other logical
as-is data
architecture
components

As-is

To-be

Technology
Independent/
Logical

abstraction

Technology
Dependent/
Physical

Data Models Used to Support Strategy

- Flexible, adaptable data structures
- Cleaner, less complex code
- Ensure strategy effectiveness measurement
- Build in future capabilities
- Form/assess merger and acquisitions strategies
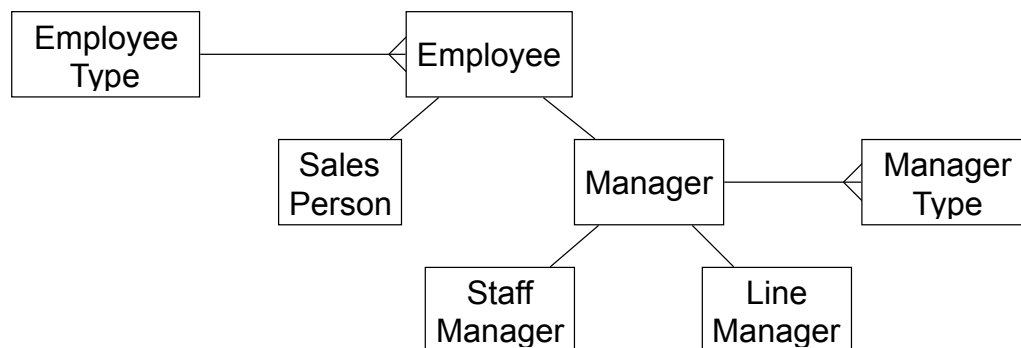
Adapted from Clive Finkelstein *Information Engineering Strategic Systems Development* 1992

How do Data Models Support Organizational Strategy?

- Consider the opposite question:

  – Were your systems explicitly designed to be integrated or otherwise work together?

  – If not then what is the likelihood that they will work well together?

  – In all likelihood your organization is spending between 20-40% of its IT budget compensating for poor data structure integration

  – They cannot be helpful as long as their structure is unknown

- Two answers

  – Achieving efficiency and effectiveness goals

 – Providing organizational dexterity for rapid implementation

Data Modeling Ensures Interoperability
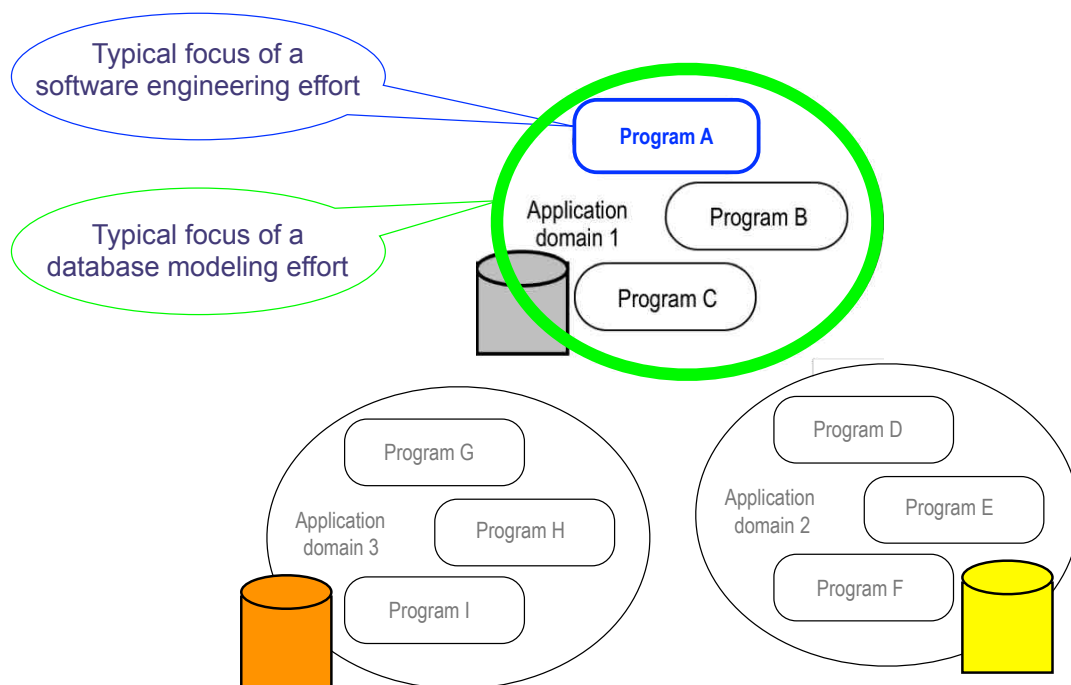
Data Model Focus has Great Potential Business Value

- How are decisions about the range and scope of common data usage, made?

- Analysis scope is on use of data to support a process

- Problems caused by data exchange or interface problems

- Goals often connect strategic and operational

- One data model is ideal
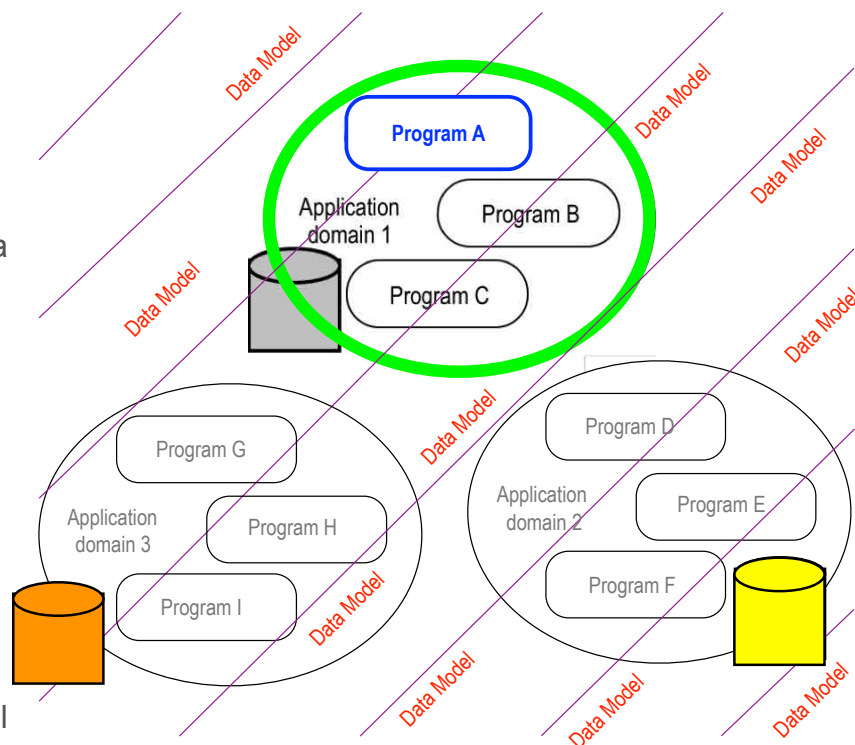
Data Modeling Fundamentals

- Data Management Contextual Overview
- Motivation
  - of systems/components
  - Data is not well understood
- Why data modeling & what is it?
  - Model represents our understanding of the
  - Fundamental, foundational system characteristics
  - Shared between system and human
- Fundamentals
  - The power of the purpose statement
  - Understanding data centric thinking
  - Data modeling compliments other architecture/ engineering techniques, as well as
  - Challenges beyond data modeling
- Take Aways, References & Q&A

Use Models to

- Store and formalize information

- Filter out extraneous detail

- Define an essential set of information

- Help understand complex system behavior

- Gain information from the process of developing and interacting with the model

- Evaluate various scenarios or other outcomes indicated by the model

- Monitor and predict system responses to changing environmental conditions

---

Data Modeling for Business Value

- Goal must be shared IT/business understanding
  - No disagreements = insufficient communication
- Data sharing/exchange is largely and highly automated and thus dependent on successful engineering
  - It is critical to engineer a sound foundation of data modeling basics (the essence) on which to build advantageous data technologies
- Modeling characteristics change over the course of analysis
  - Different model instances may be useful to different analytical problems
- Incorporate motivation (purpose statements) in all modeling
  - Modeling is a problem defining as well as a problem solving activity - both are inherent to architecture
- Use of modeling is much more important than selection of a specific modeling method
- Models are often living documents
  - It easily adapts to change
- Models must have modern access/interface/search technologies
  - Models need to be available in an easily searchable manner
- Utility is paramount
  - Adding color and diagramming objects customizes models and allows for a more engaging and enjoyable user review process

Inspired by: Karen Lopez http://www.information-management.com/newsletters/enterprise_architecture_data_model_ERP_BI-10020246-1.html?pg=2

Upcoming Events

August Webinar
Data Management versus Data Strategy
August 13, 2019 @ 2:00 PM ET (UTC-4)

September Webinar
Getting Started with Data Stewardship
September 10, 2019 @ 2:00 PM ET (UTC-4)

Sign up for webinars at:
www.datablueprint.com/webinar-schedule

Mark your calendar for our monthly #DataEd webinars

Brought to you by **DATAVERSITY**

Questions?

Ask Questions

It's your turn!
Use the chat feature or
Twitter (#dataed) to submit
your questions now!

data blueprint

UNLOCKING BUSINESS VALUE

datablueprint.com

10124 W. Broad Street, Suite C
Glen Allen, Virginia 23060
804.521.4056