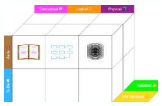
Conceptual

Versus



Logical

Versus

Physical

Data Modeling

peter.aiken@anythingawesome.com +1.804.382.5957

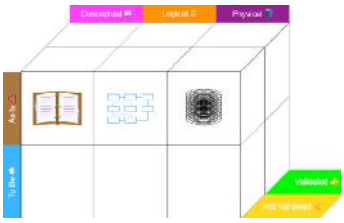
Peter Aiken, Ph.D.

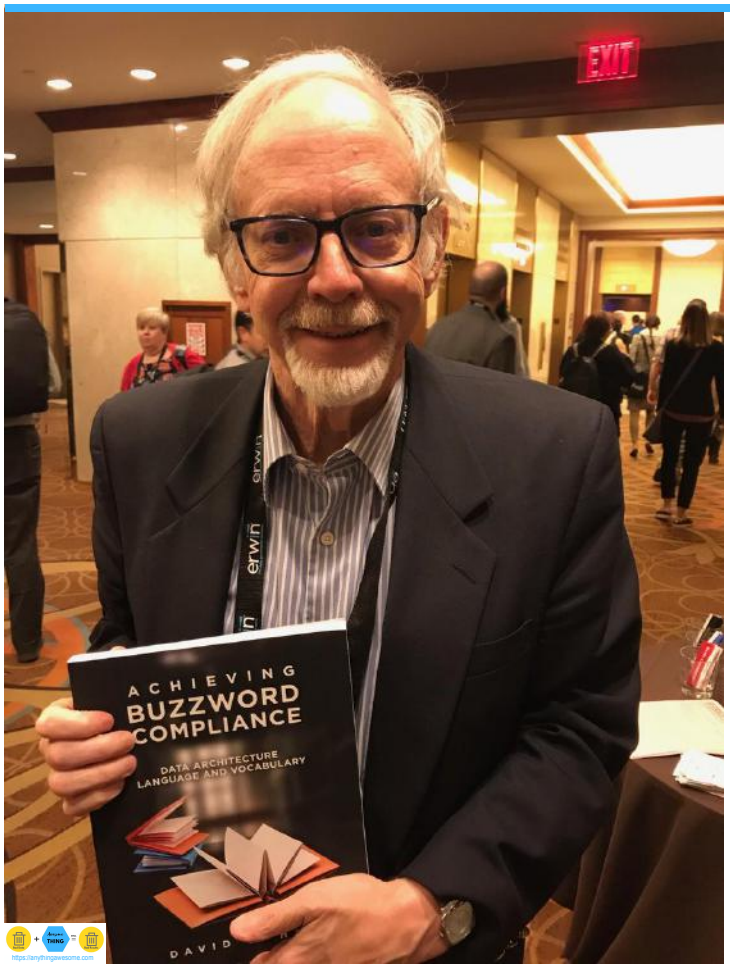
- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- Multi-year immersions
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - HUD …
- 12 books and dozens of articles

Program Overview

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

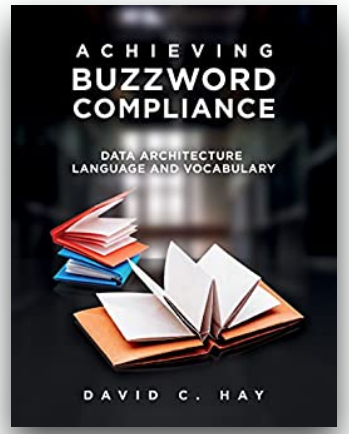
- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A





Achieving Buzzword Compliance

Data Architecture Language and Vocabulary



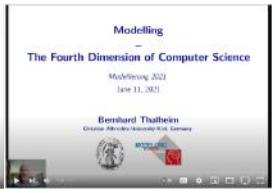
amazon.com link:

https://www.amazon.com/Achieving-Buzzword-Compliance-Architecture-Vocabulary-ebook/dp/B07FG1WRSD/ref=sr_1_1?crid=2QL3ZWKU2L3VC&keywords=Achieving+Buzzword+Compliance%3A+Data+Architecture+Language+and+Vocabulary&qid=1657032460&sprefix=achieving+buzzword+compliance+data+architecture+language+and+vocabulary%2Caps%2C324&sr=8-1

Research Efforts

- Professor Bernhard Thalheim and associated research efforts have contributed much to these topics including:

- Conceptual modelling
  - https://www.youtube.com/watch?v=Y9_7KSsSUpg
  - https://www.youtube.com/watch?v=mKcwbR6uJwU



- Claim: logical models also conceptual models
  - https://www.youtube.com/watch?v=L8yGjEbwTsQ
  - https://link.springer.com/article/10.1007/s10270-020-00836-z
  - https://dl.acm.org/doi/10.1007/s10270-020-00836-z

Https://Www.Youtube.Com/Watch?V=9qWjpVtr_Hg&T=2s

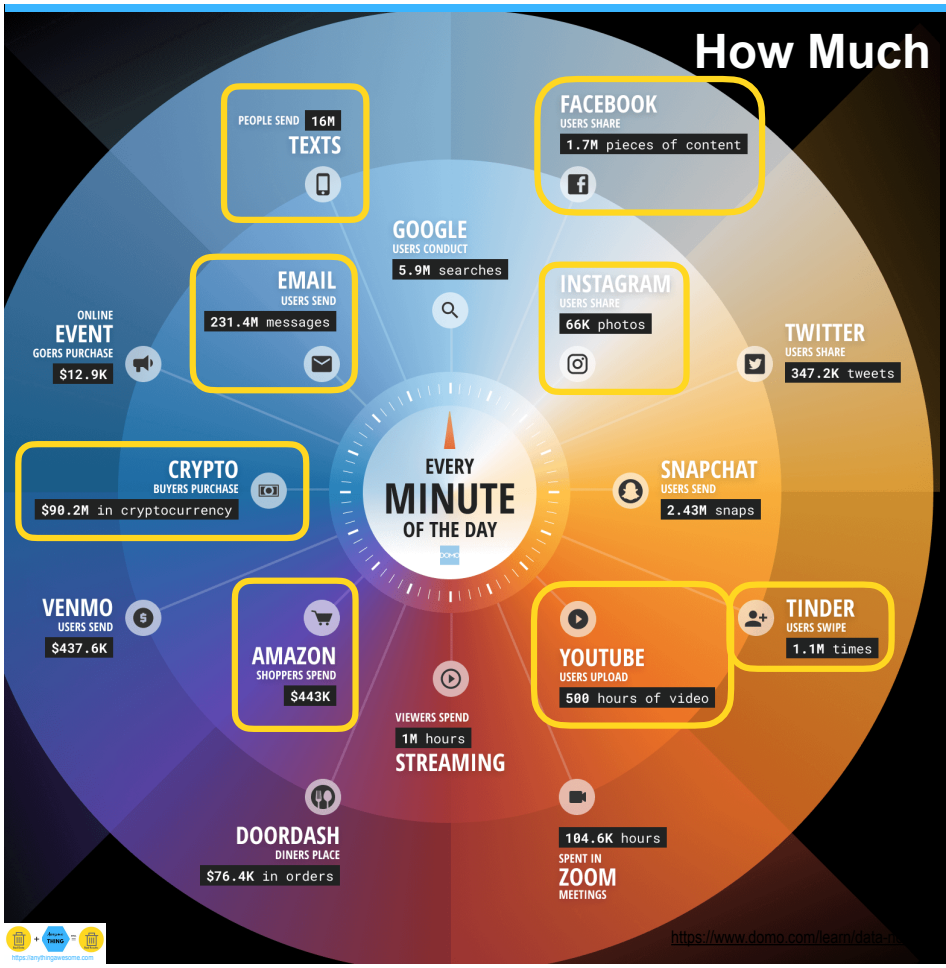
Advanced Data Modeling class 2016

Conceptual vs. Logical vs. Physical

Stages of Data Modeling

©Gordon C. Everest
Professor Emeritus of MIS and Database
Carlson School of Management
University of Minnesota
geverest@umn.edu http://geverest.umn.edu

How Much Data (by the Minute?)

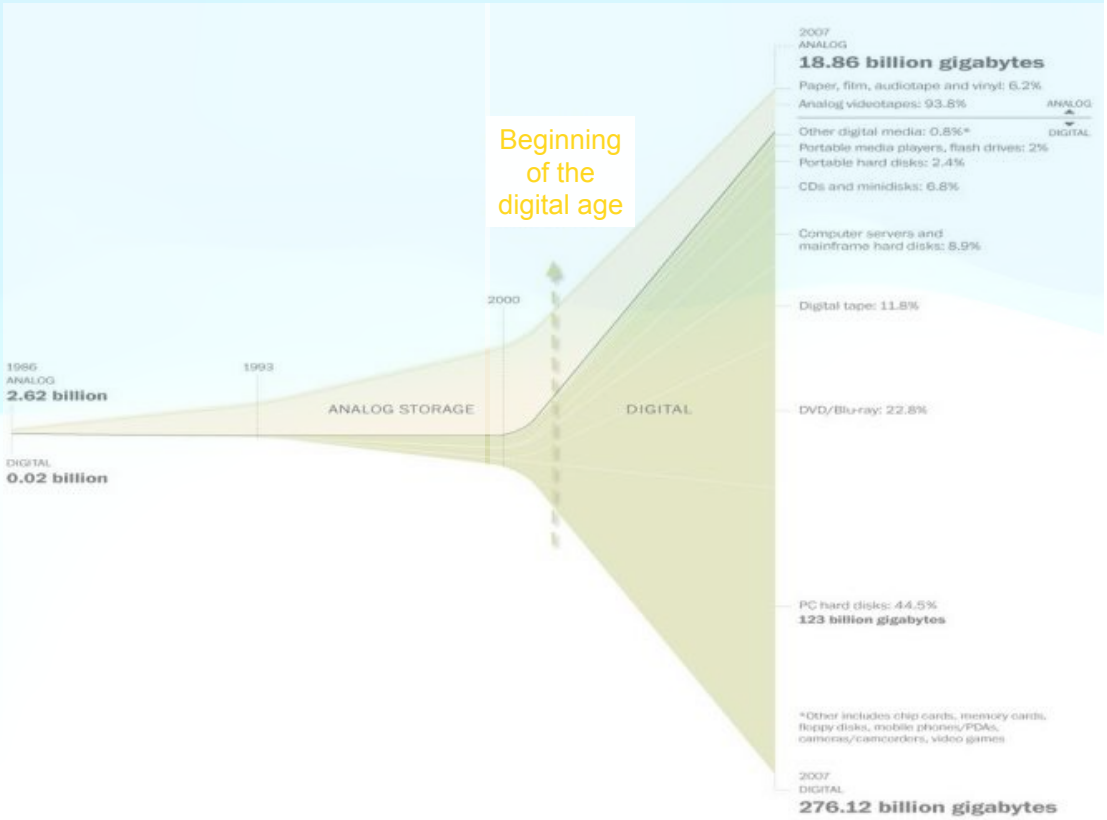


For the entirety of 2022, every minute of every day:

- Facebook users share 1.7 million pieces of content
- Instagram users share 66K photos
- Tinder users record 1M swipes
- YouTube users upload 500 hours of video
- Amazon shoppers spend $443K
- Crypto buyers purchase $90M+
- Email users send 231M messages
- People send 16M texts

https://www.domo.com/learn/data-i...

Global Information Storage Capacity

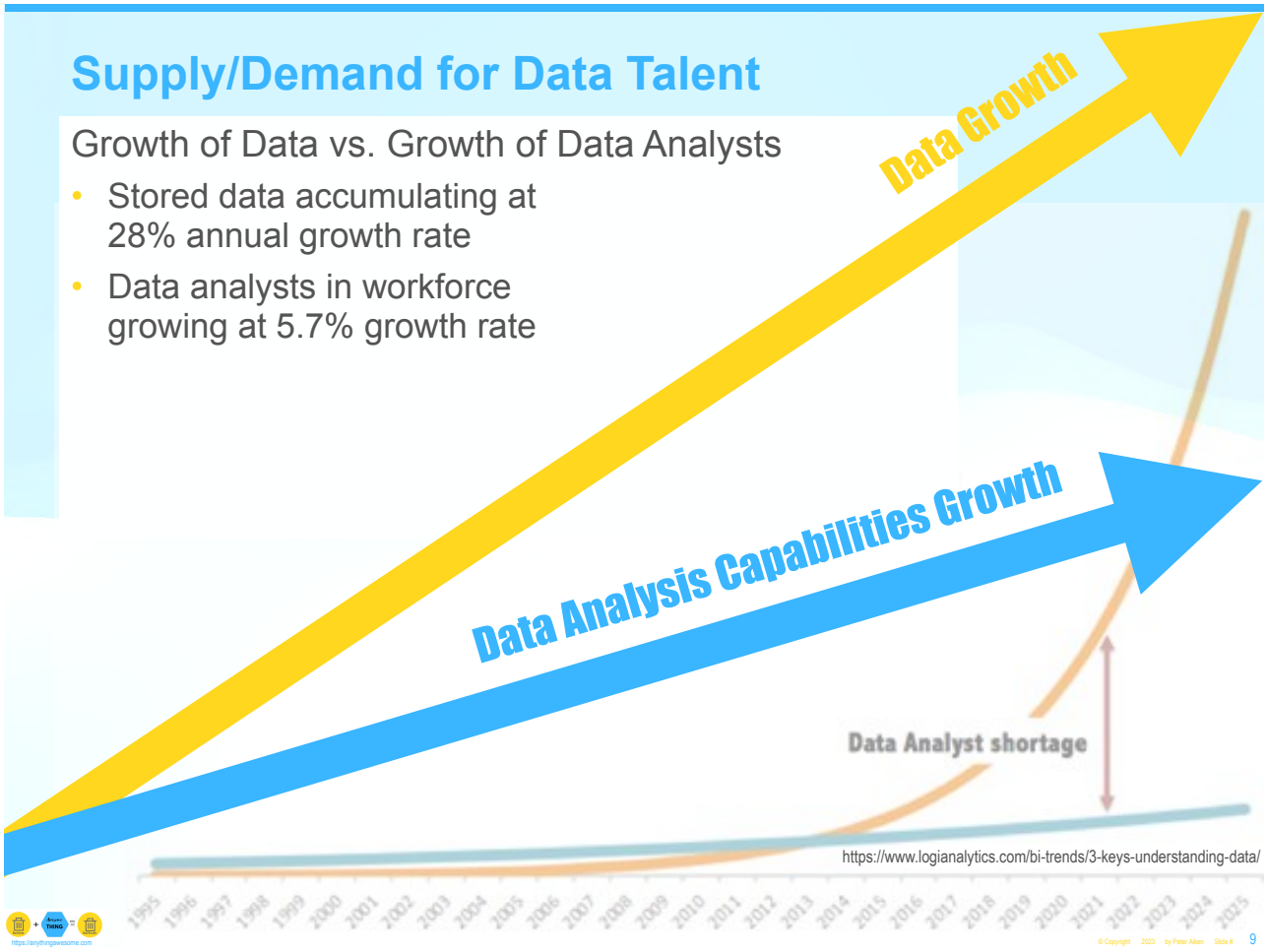


https://www.martinhilbert.net/worldinfocapacity-html/

Supply/Demand for Data Talent

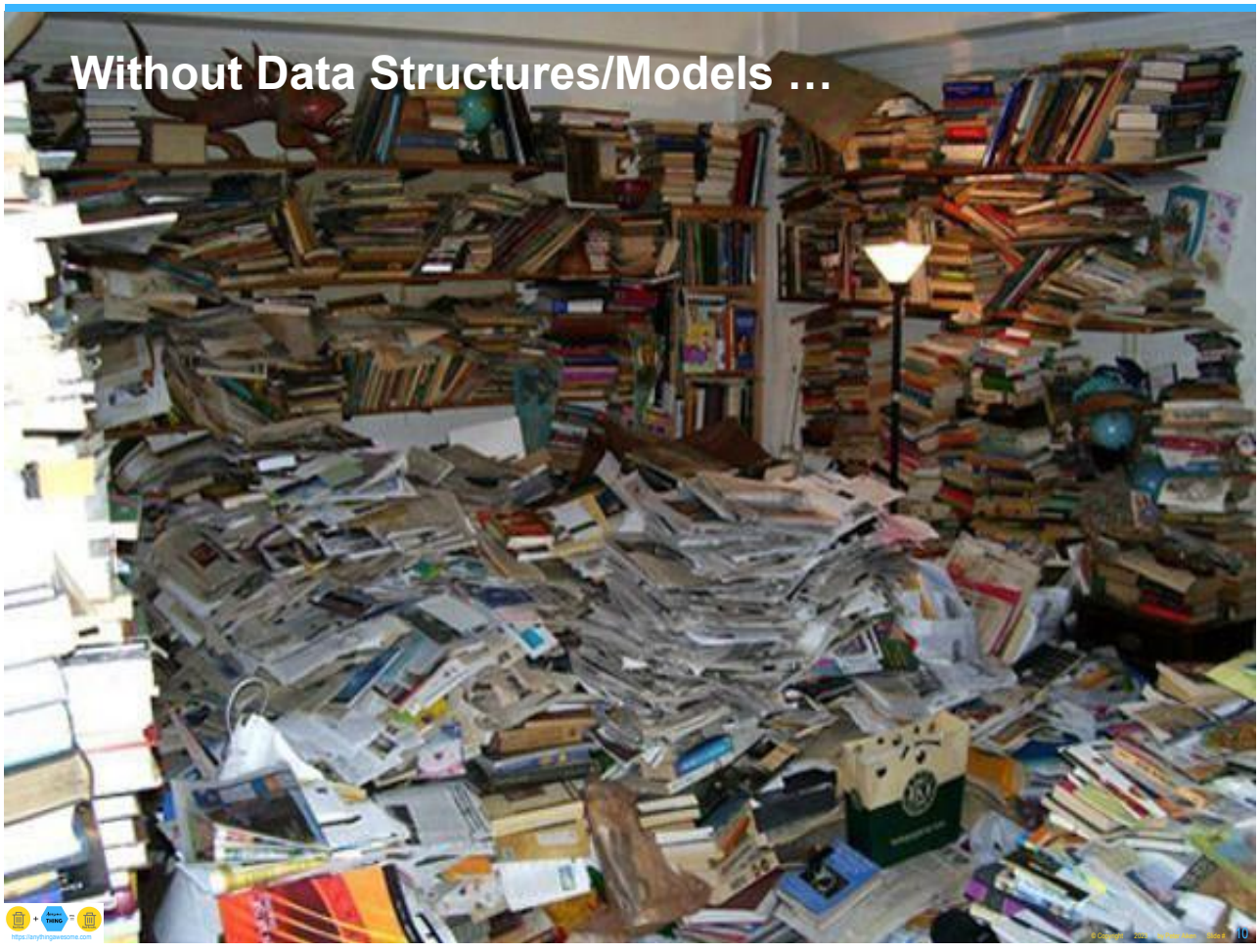
Growth of Data vs. Growth of Data Analysts

- Stored data accumulating at 28% annual growth rate
- Data analysts in workforce growing at 5.7% growth rate



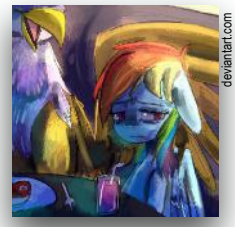
https://www.logianalytics.com/bi-trends/3-keys-understanding-data/

Without Data Structures/Models …

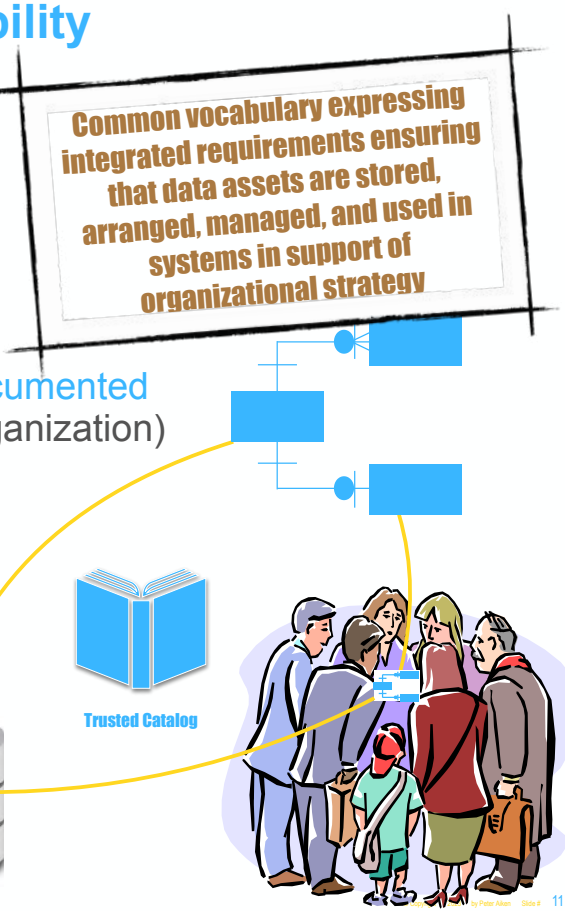


Understanding = Interoperability

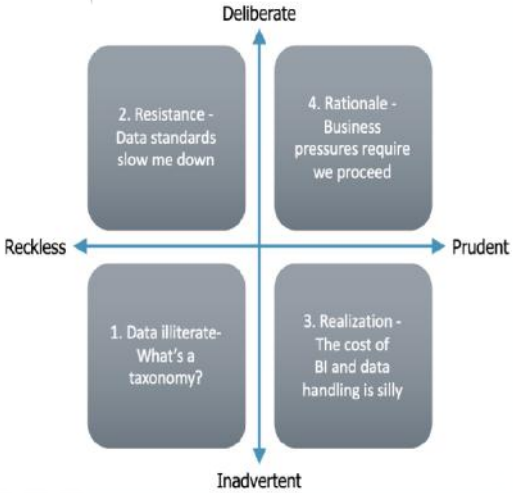
- All organizations have architectures
  - Business
  - Process
  - Systems
  - Security
  - Technical
  - Data/Information



- Some are better understood and documented (and therefore more useful to the organization)
- '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
  - Business
  - Technical
  - Systems



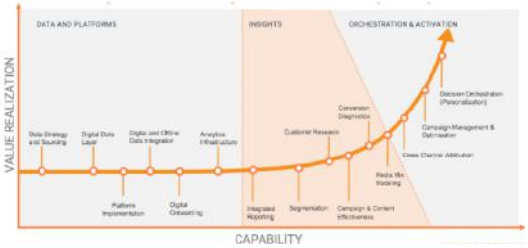
Modeling Addresses Data Debt Proactively



https://johnladley.com/a-bit-more-on-data-debt/

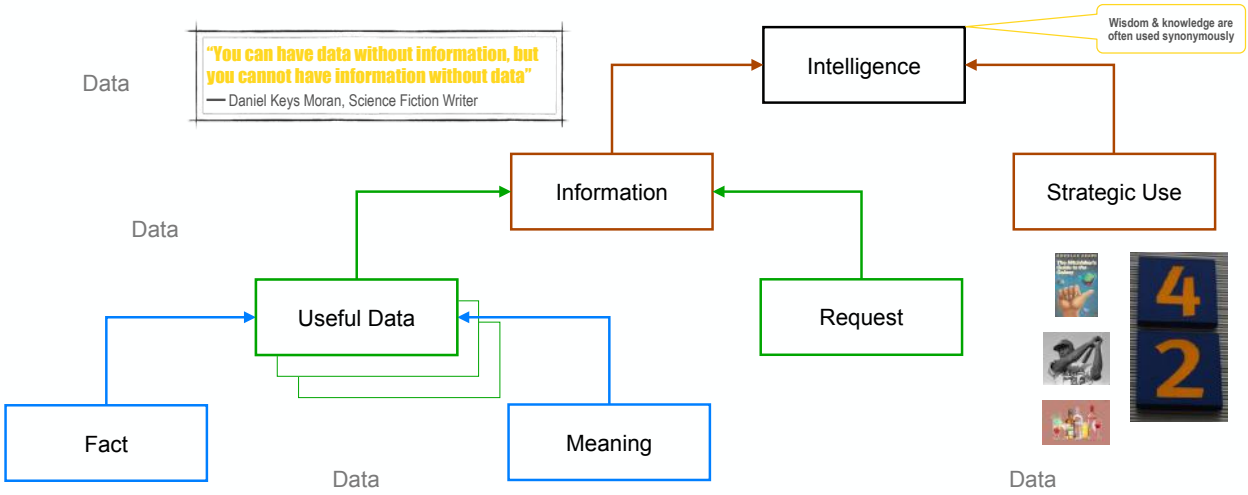
- Data debt
  - The time and effort it will take to return your shared data to a governed state from its (likely) current state of ungoverned
- Getting back to zero
  - Involves undoing existing stuff
  - Likely new skills are required

https://uk.nttdataservices.com/en/blog/2020/february/how-to-get-rid-of-your-data-debt



https://www.merkleinc.com/blog/are-you-buried-alive-data-debt

A Model Precisely Defining 3 Important Concepts



1. Each FACT combines with one or more MEANINGS.
2. Each specific FACT and MEANING combination is referred to as a DATUM.
3. An INFORMATION is one or more DATA that are returned in response to a specific REQUEST
4. INFORMATION REUSE is enabled when one FACT is combined with more than one MEANING.
5. INTELLIGENCE is INFORMATION associated with its STRATEGIC USES.
6. DATA/INFORMATION must formally arranged into an ARCHITECTURE.

[Built on definitions from Dan Appleton. 1983]

Each Data Arrangement Is a Data Structure

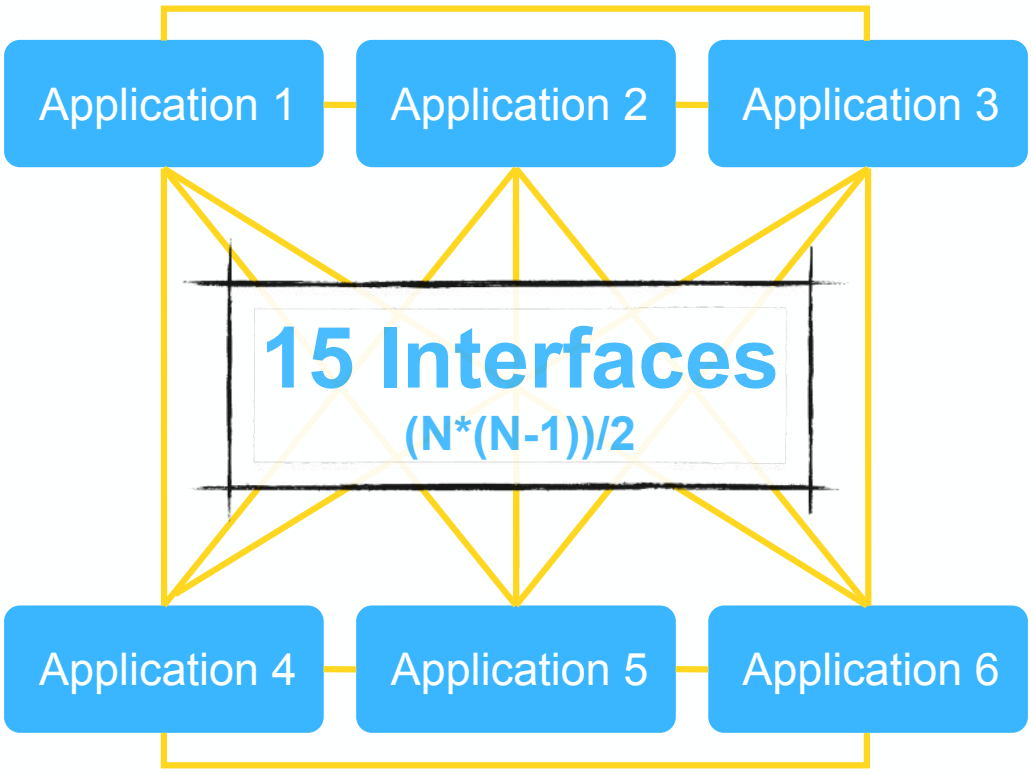
"*An organization of information, usually in memory, for better algorithm efficiency, such as queue, stack, linked list, heap, dictionary, and tree, or conceptual unity, such as the name and address of a person. It may include redundant information, such as length of the list or number of nodes in a subtree.*"

Some data structure characteristics

- Grammar for data objects
  - Grammar is the principles or rules of an art, science, or technique "a grammar of the theater"
- Data Object Constraints
- Ordering
  - Sequential, hierarchical, relational, network, lake, other
- Uniqueness
- Balance
- Optimality
- Future enhanceability
  - Multi-currency
  - Device handoff features

http://www.nist.gov/dads/HTML/datastructur.html

How Many Interfaces Are Required To Solve This Integration Problem?

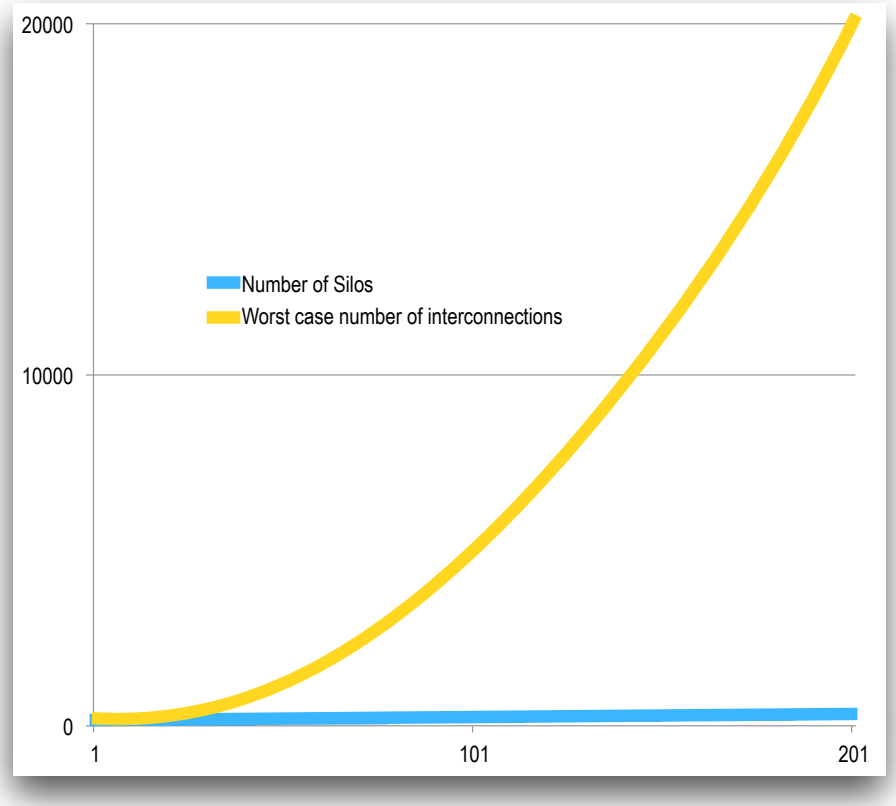


RBC: 200 applications - 4900 batch interfaces

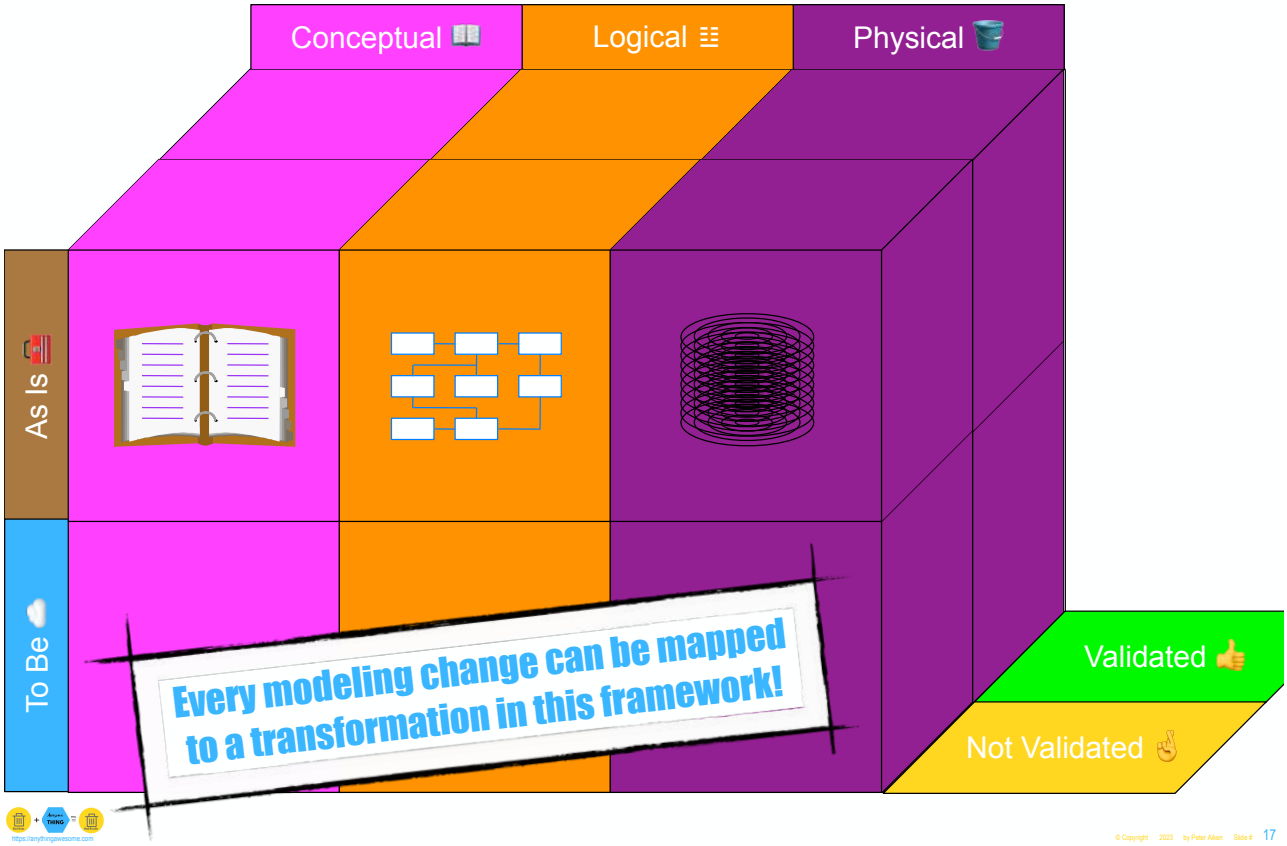
The Rapidly Increasing Cost of Complexity

N

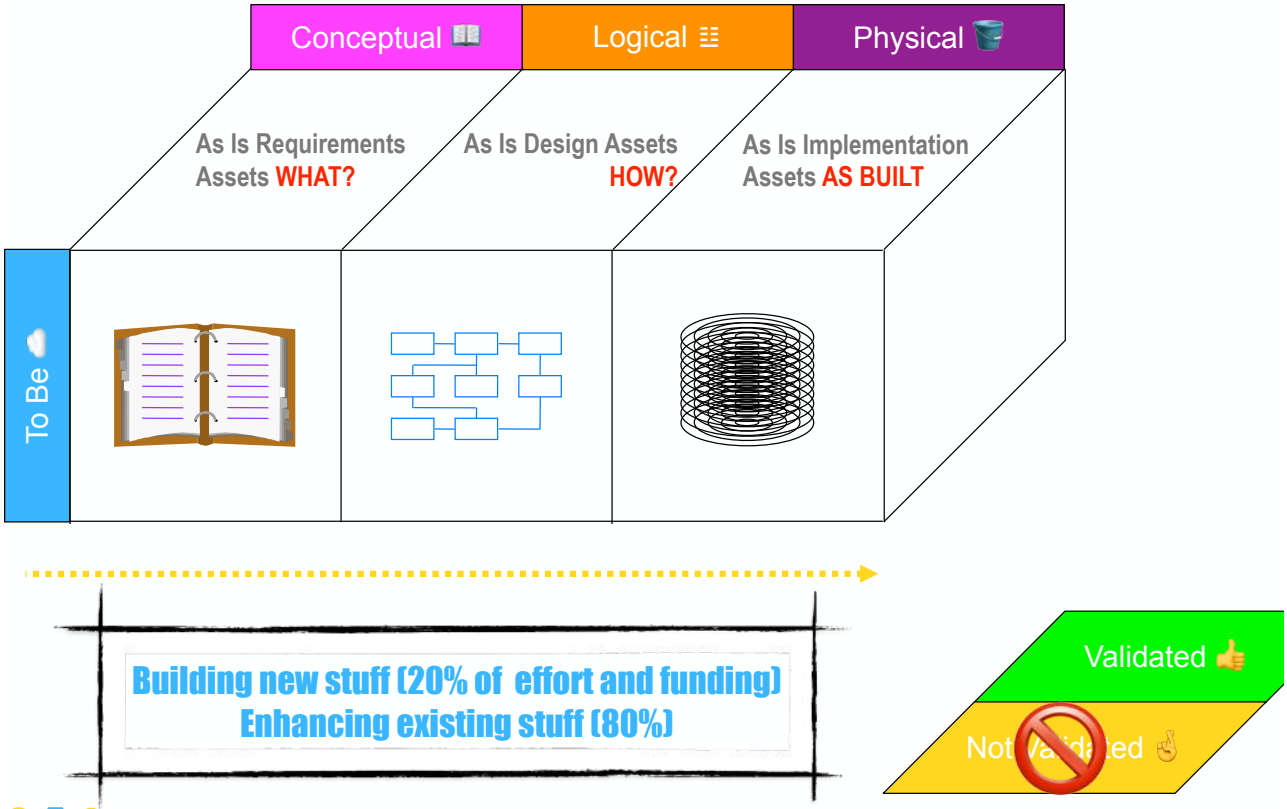
- 6 / 15
- 60 / 1,770
- 600 / 179,700
- 200 / 19,900
- 200 / 5,000 (actual)



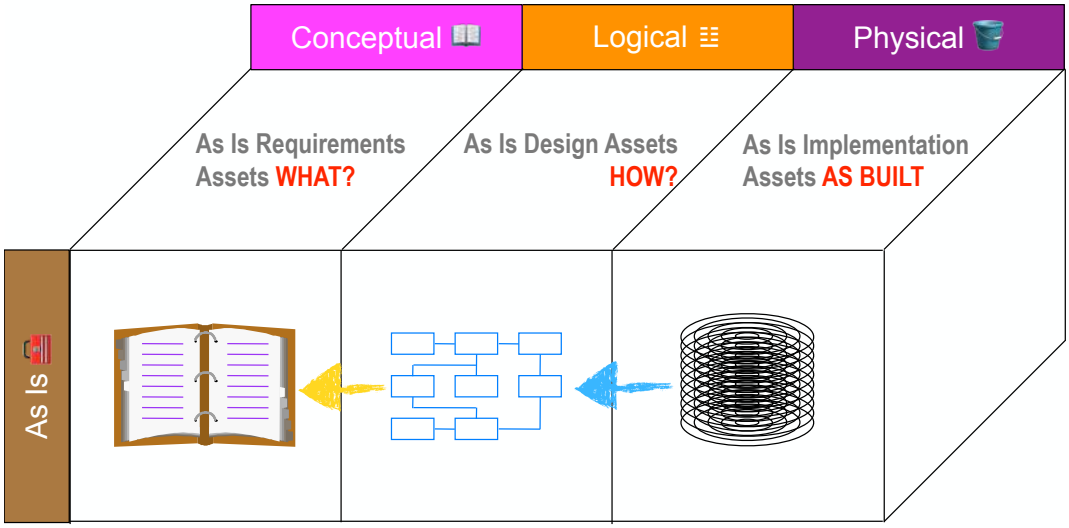
3-Dimensional Model Evolution Framework



Forward Engineering

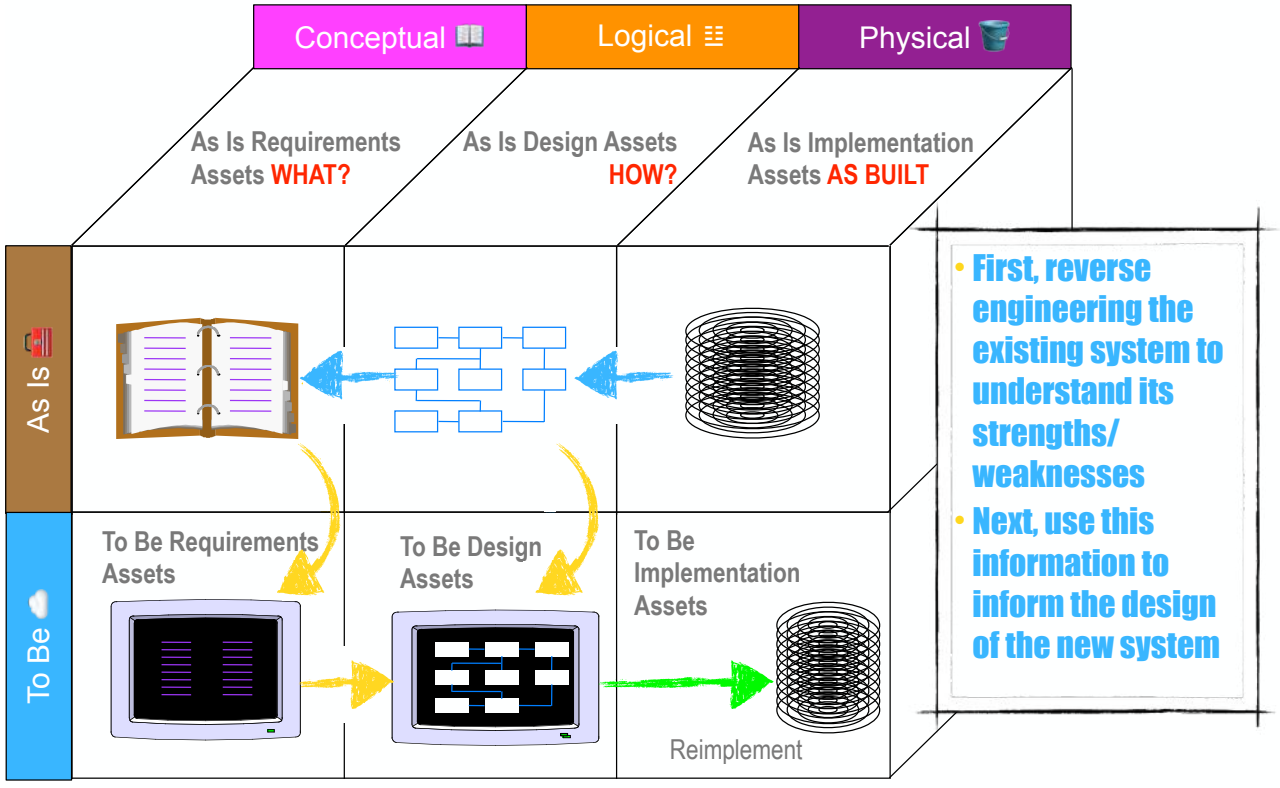


80% Reverse Engineering

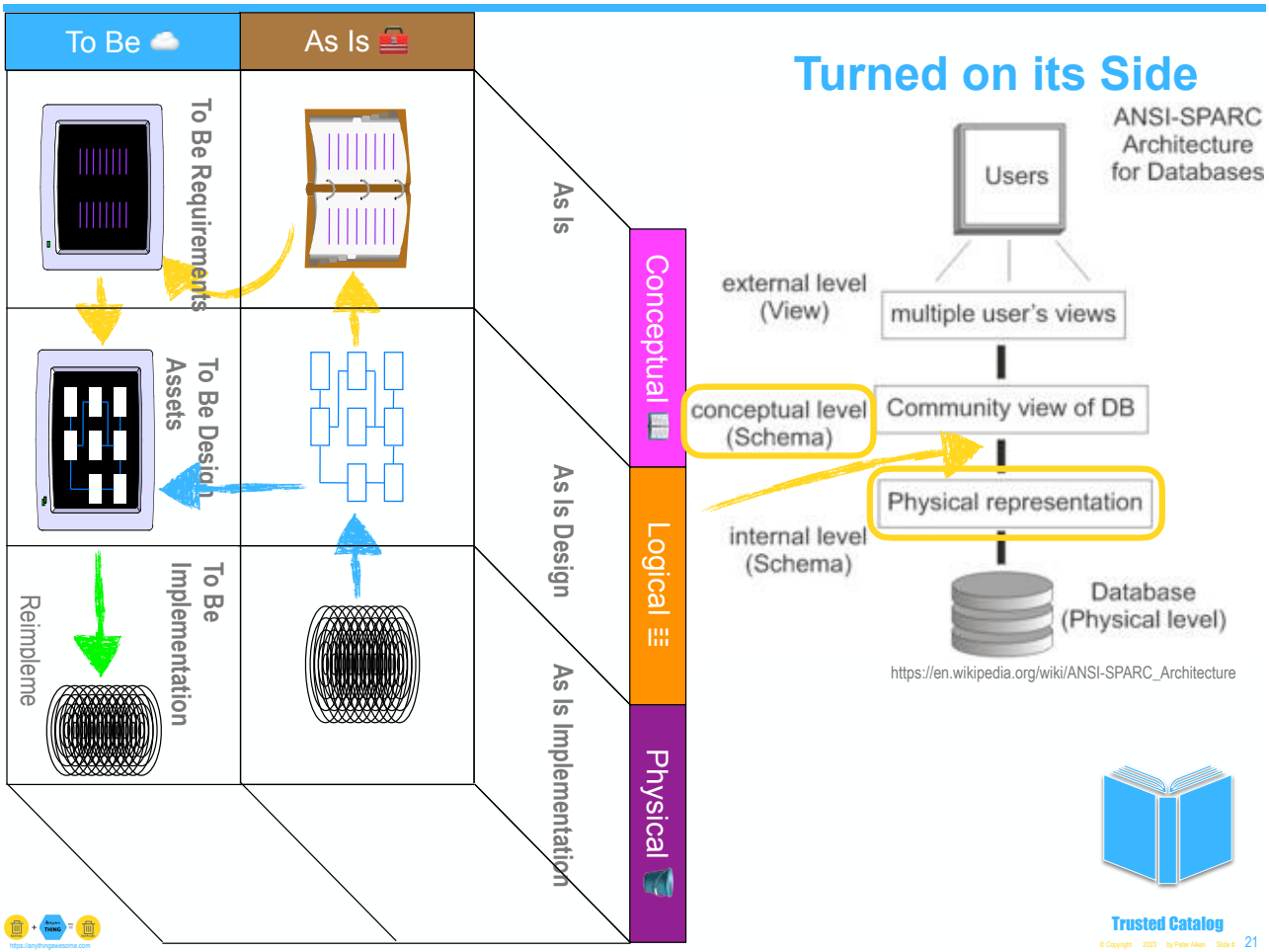


Evolve existing systems using a structured technique aimed at recovering rigorous knowledge of the existing system to leverage enhancement efforts [Chikofsky & Cross 1990]

Reengineering

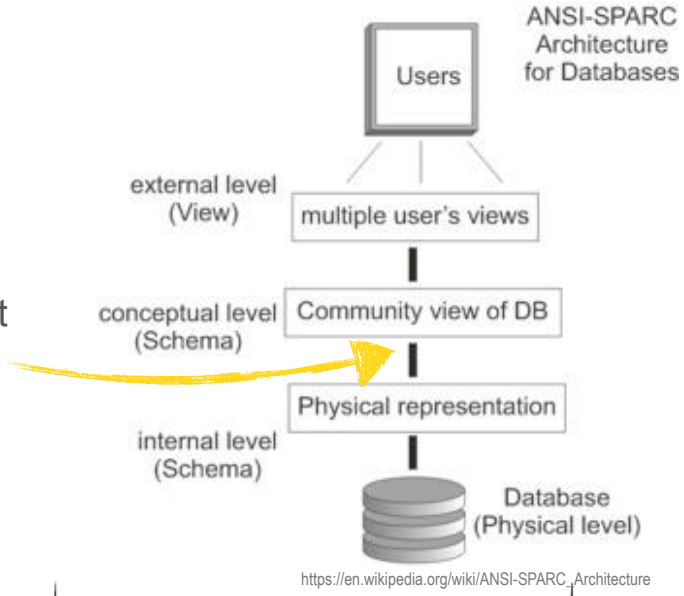


- **First, reverse engineering the existing system to understand its strengths/ weaknesses**
- **Next, use this information to inform the design of the new system**



ANSI-SPARC 3-Layer Schema

- **Conceptual** - Highest level of abstraction, focused on data requirements (what), linked directly to strategy
- **Logical** - Usually a refinement of conceptual model, focused on how data requirements are met using business terminology
- **Physical** - Implementation of the logical model with security, configuration management, and implementation specific details, specified via DDL

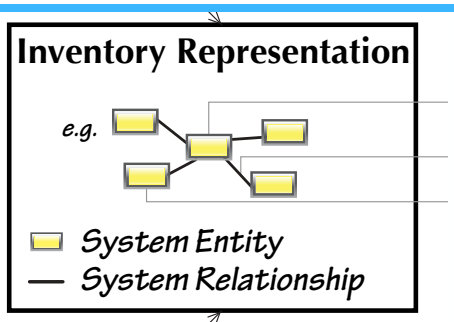


When changing to a new DBMS technology, the database administrator should be able to change the conceptual or global structure of the database without affecting the users

Zachman Framework

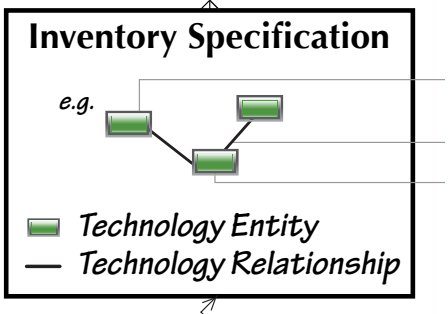
Rows 3, 4, & 5 of the "What" Column

Architect Perspective (Business Logic Designers)



Conceptual

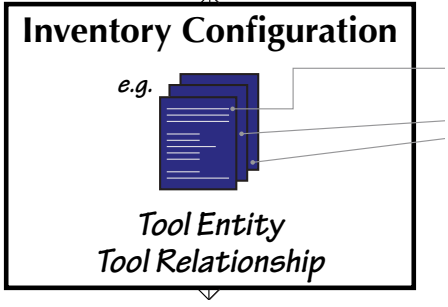
Engineer Perspective (Business Physics Builders)



Logical

Technician Perspective (Business Component Implementers)

Alignment Tr



Physical

https://www.britannica.com/video/179912/Overview-Millau-Viaduct-France-Tarn-River

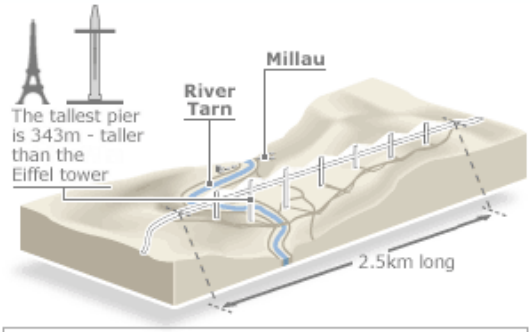
Millau Viaduct Is the Highest Bridge in the World as of 2007

Conceptual Models

- Business focused
- Entity level
- Provides focus, scope, and guidance to modeling effort
- Sometimes thrown away - rarely maintained

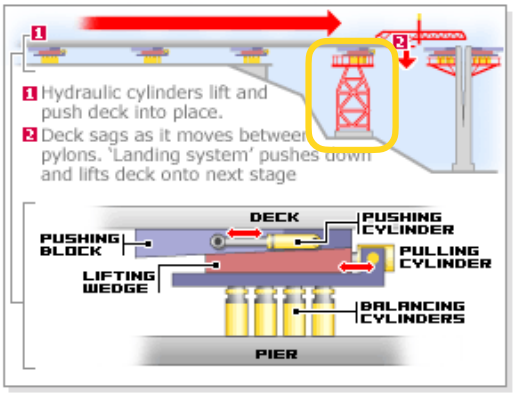
Logical Models

- Required to achieve the transition from conceptual to physical
- Developed to the attribute level and understood at 3rd normal form
- Logical models are developed to be refined to until it becomes a solution - sometimes purchased (as in EDW) always requires tailoring
- Used to guarantee the rigor of the data structures by formally describing the relationship between data items in a strong fashion
- More often maintained

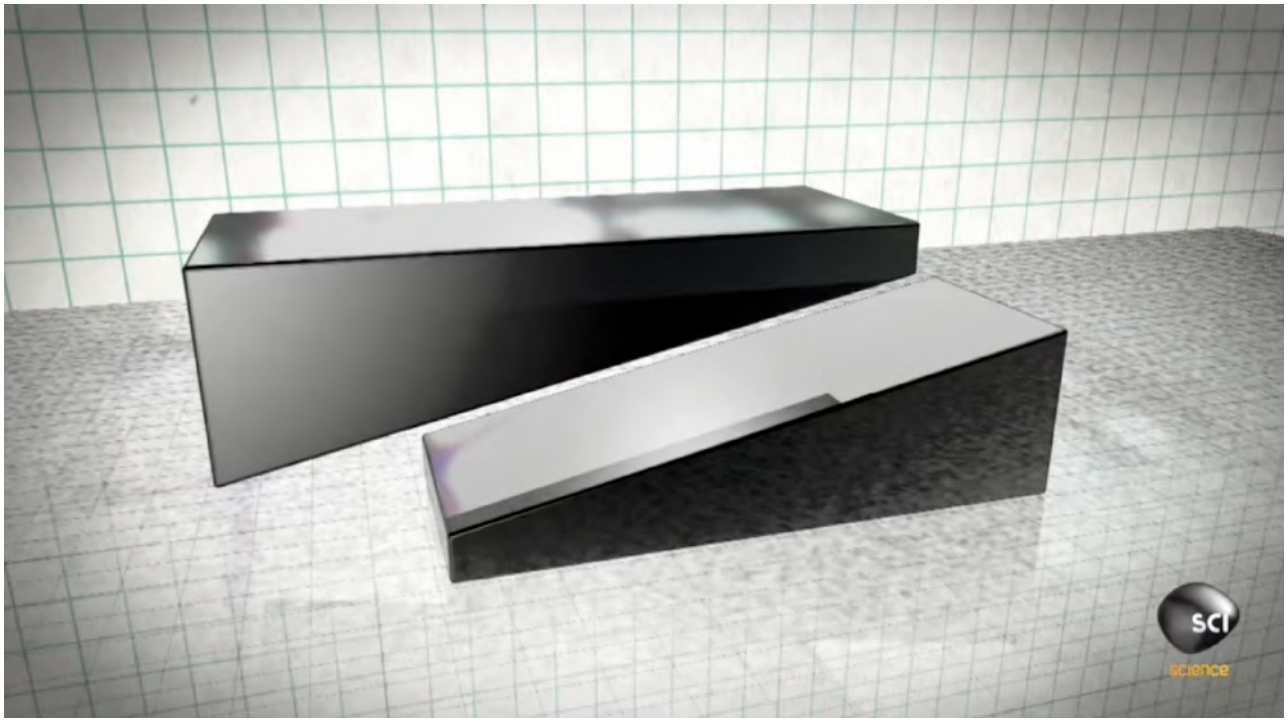


Physical Models

- Become the blueprints for physical construction of the solution
- Blueprints are used for future maintenance of the solution



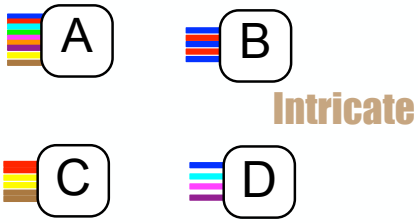
Avoiding any Side-Pressure on the Supporting Piers

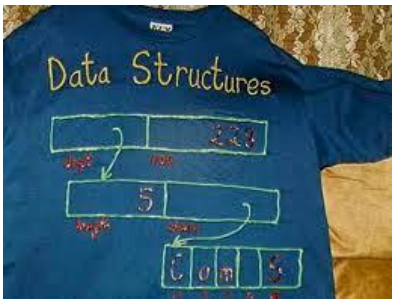


https://www.youtube.com/watch?v=_iK0soIvjv8 & https://www.youtube.com/watch?v=DlbTNJ0AU1Y

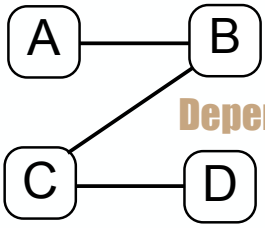
How Are Components Expressed as Architectures?

- Details are organized into larger components

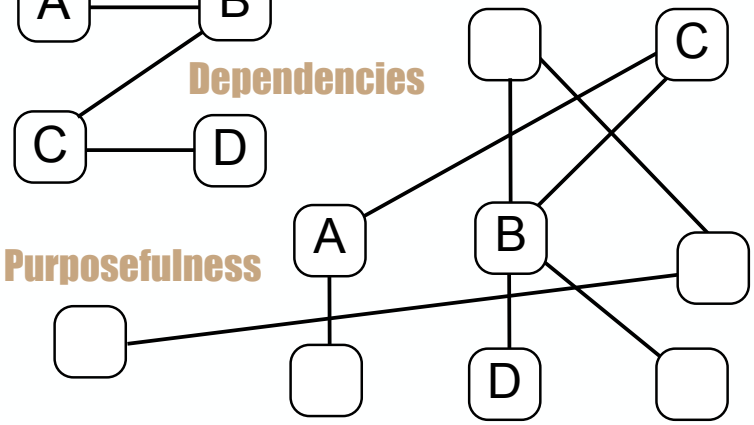




- Larger components are organized into models



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



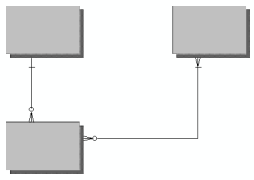
How Are Data Structures Expressed as Architectures?

- Attributes are organized into entities/objects
  - Attributes are characteristics of "things"
  - Entitles/objects are "things" whose information is managed in support of strategy
  - Example(s)

Intricate

- Entities/objects are organized into models
  - Combinations of attributes and entities are structured to represent information requirements
  - Poorly structured data, constrains organizational information delivery capabilities
  - Example(s)

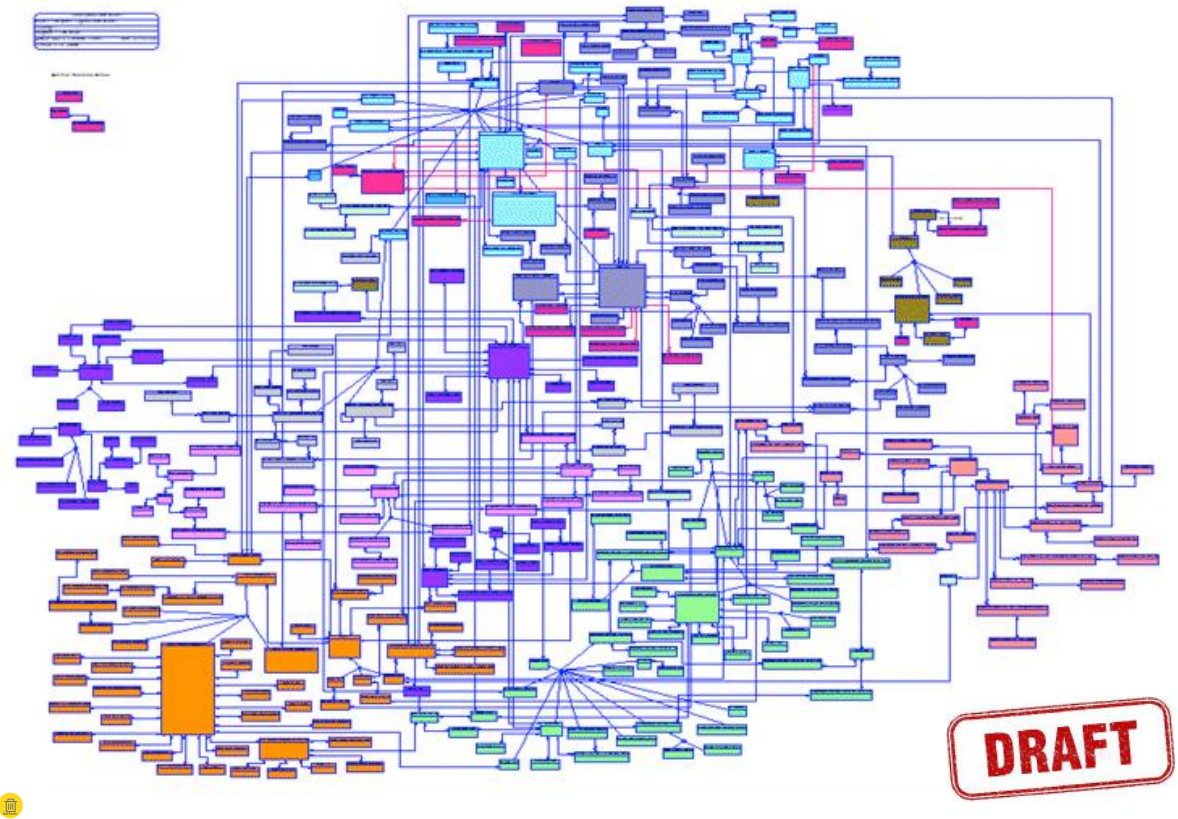
Dependencies



- Models are organized into architectures
  - 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
  - *Why no examples?*

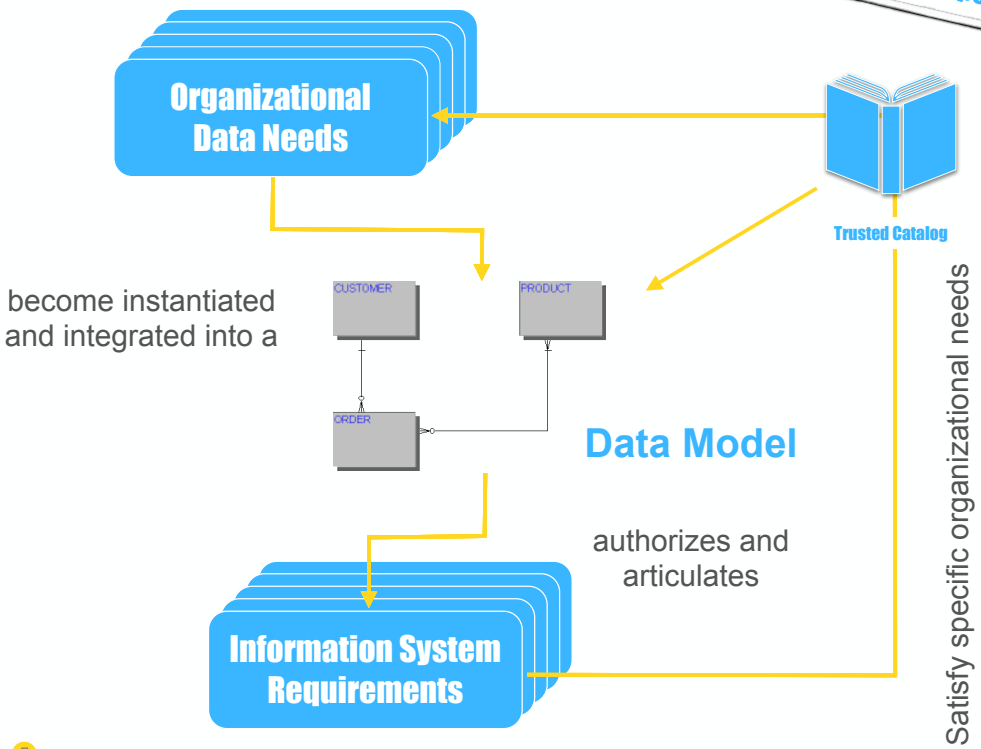
Purposefulness

Data Architectures Are Composed of Data Models



Data Modeling Is Iterative

Data models and data architectures are developed in response to needs

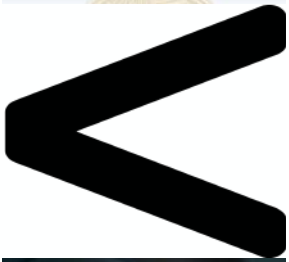


The Princess on the Pea

by
Hans Christian Andersen

Doing a Poor Job With Data Modeling

- Failure to understand the role of data governance re: proposed and existing software/services
  - Locks in imperfections for the life of the application
  - Restricts data investment benefits
  - Decreases organizational data leverage
- Accounts for 20-40% of IT budgets devoted to evolving
  - Data migration (Changing the data location)
  - Data conversion (Changing data form, state, or product)
  - Data improving (Inspecting and manipulating, or re-keying data to prepare it for subsequent use)
- Lack of data governance causes everything else to
  - Take longer
  - Cost more
  - Deliver less
  - Present greater risk (with thanks to Tom DeMarco)



(A Hypothetical Portion of the) iTunes → Music™ Database

- Question:
  - What information is lost if we delete record #1?

Row	Purchaser ID	Song	Price
1	Peter	We Met Today	$0.99
2	Peter	My Mother's Voice	$1.29
3	Peter	Fortune Smiles	$0.99
4	Lolly	Thousand Pieces of Gold	$0.99

(A Hypothetical Portion of the) Music™ Database: Deletion Anomaly

- Question:
  - What information is lost if we delete record #1?
- Answer:
  - We loose the fact that Peter purchased "We Met Today"
  - We also loose the fact that "We Met Today" costs $0.99
  - These are usually undesirable and unintended

Row	Purchaser ID	Song	Price
1	Peter	We Met Today	$0.99
2	Peter	My Mother's Voice	$1.29
3	Peter	Fortune Smiles	$0.99
4	Lolly	Thousand Pieces of Gold	$0.99

Music™ Database: Insertion Anomalies

- Question:
  - Suppose we want to add new song SCUBA and that it costs $1.29?
- Answer:
  - Cannot enter it until a purchaser buys SCUBA
  - We cannot insert a full row until we have an additional fact about that row
  - This is usually undesirable and unintended

Row	Purchaser ID	Song	Price
1	Peter	We Met Today	$0.99
2	Peter	My Mother's Voice	$1.29
3	Peter	Fortune Smiles	$0.99
4	Lolly	Thousand Pieces of Gold	$0.99
5	???	SCUBA	$1.29

Music™ Database: Update Anomalies

- Question:
  - Suppose we want to increase the price of 'We Met Today' from $0.99 to $1.29?
- Answer:
  - Change to data items such as Song requires examination of every single record
  - Will not catch spelling errors - such as "We met Toddy"
  - This is usually undesirable and unintended

Row	Purchaser ID	Song	Price
1	Peter	We Met Todday	$0.99
2	Peter	My Mother's Voice	$1.29
3	Peter	Fortune Smiles	$0.99
4	Lolly	Thousand Pieces of Gold	$0.99
5	???	SCUBA	$1.29

There Are Correct Ways To Organize Data

- Optimization can be done for:
  - Flexibility
  - Adaptability
  - Retrievability
  - Risk reduction
  - ...
- Techniques include:
  - Data integrity
  - Smart codes bad/dumb codes good
  - Architecture (table joins)
  - ...

ORIGINAL

Record	Purchaser ID	Song	Pric
1	Purchaser #1	Cool Walk (Live)	$1.99
2	Purchaser #1	Sushi (Live)	$0.99
3	Purchaser #1	Love Ballade (Live)	$0.99
4	Purchaser #2	A Salute to Bach	$0.99
5	Purchaser #3	Coolwalk (Live)	$1.99

How Should It Be Done? (In General)

- As much as possible, store 1 fact per row
  - Row 2 is a good example as it shows both that Purchaser #1 has purchased Sushi (Live) and that it costs $0.99
  - These are two distinct facts and are correctly stored in two tables sharing a formal relationship
  - More remains coded

ORIGINAL

Record	Purchaser ID	Song	Pric
1	Purchaser #1	Cool Walk (Live)	$1.99
2	Purchaser #1	Sushi (Live)	$0.99
3	Purchaser #1	Love Ballade (Live)	$0.99
4	Purchaser #2	A Salute to Bach	$0.99
5	Purchaser #3	Coolwalk (Live)	$1.99

PRICING

Record	Song	Price
1	Cool Walk (Live)	$1.99
2	Sushi (Live)	$0.99
3	Love Ballade (Live)	$0.99
4	A Salute to Bach	$0.99
5	Coolwalk (Live)	$1.99

PURCHASES

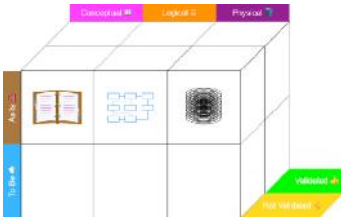
Row	Purchaser ID	Song
1	Purchaser #1	Cool Walk (Live)
2	Purchaser #1	Sushi (Live)
3	Purchaser #1	Love Ballade (Live)
4	Purchaser #2	A Salute to Bach (Medley)
5	Purchaser #3	Coolwalk (Live)
6	Purchaser #3	A Salute to Bach (Medley)

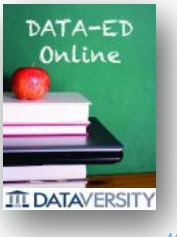
Sand in the Machinery

Program Overview

- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

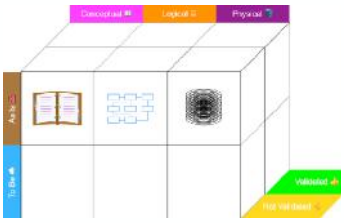




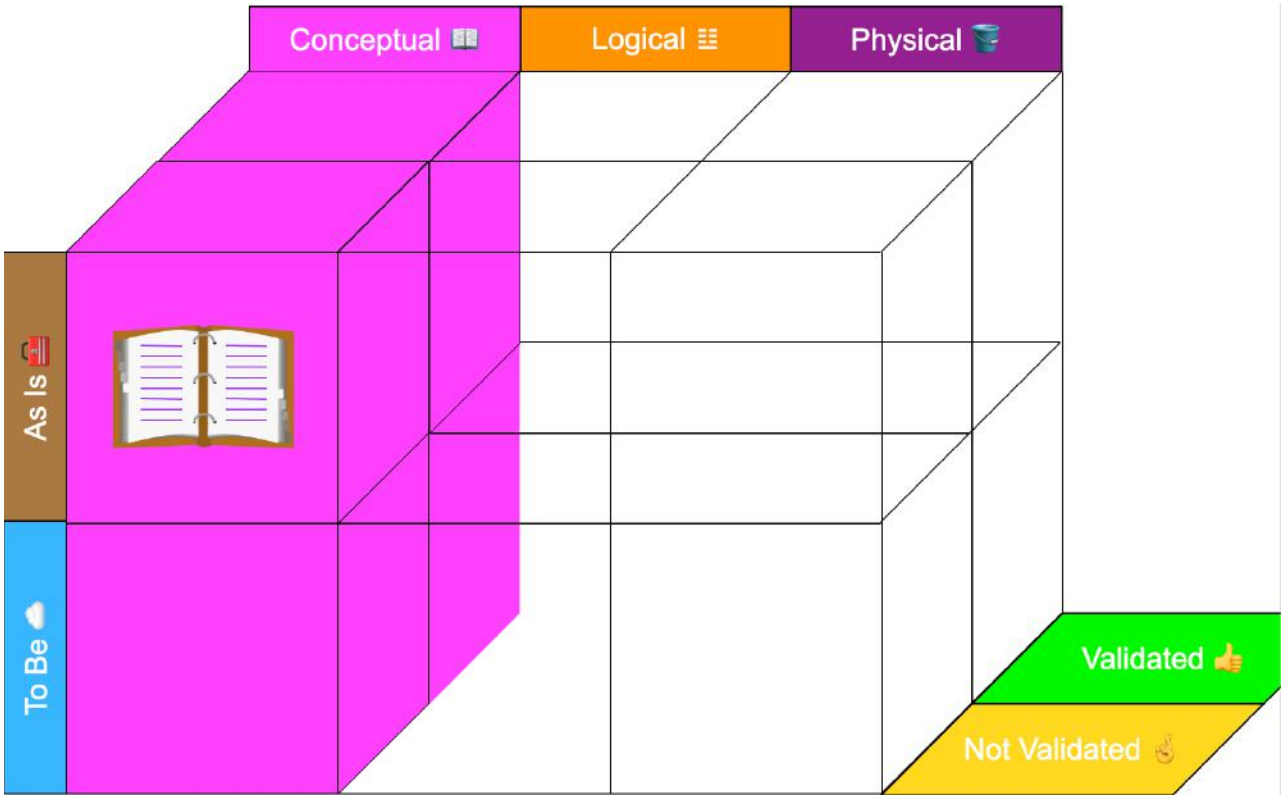
Program Overview

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

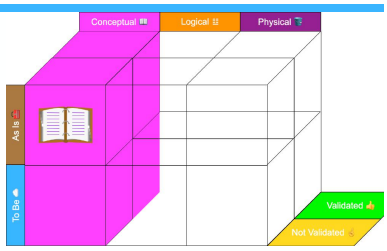
- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A



Conceptual Data Modeling



Conceptual Data Modeling



Motivation

- Harmonize/standardize vocabulary
  - Between business and technologists
  - Between humans and systems
- Focus consideration/analyses on strategic issues and tradeoffs
- Provide specifications comprising organizational data strategic objectives
- Document data requirements satisfying business objectives

Reasons for Unvalidated Conceptual Data Models

- Unvalidated models require the word on them, indicating a lack of certainty
- Useful for organizing data concepts
- Hypothesizing the relationship of various data things to various other data things

Reasons for Validated Conceptual Data Models

- Documenting the relationship of various data things to various other data things
- Standardizing on 'system-wide' definitions
- Understanding high level process interactions

Architecture Involves at Least ...

- Analysis/model evaluation
- Risk evaluation
- Volume considerations
- Workload forecasting
- Tradeoff analysis
- ...

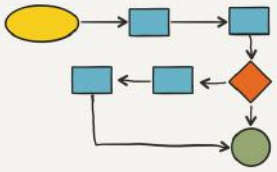
What Is Strategy?

strat·e·gy
/ˈstratəjē/

noun

1. a plan of action or policy designed to achieve a major or overall aim.
   "time to develop a coherent economic strategy"
   synonyms: master plan, grand design, game plan, plan (of action), action plan, policy, program; More

- Current use derived from military
  - **a pattern in a stream of decisions**
    [Henry Mintzberg]



PROCESS

Former Walmart Business Strategy

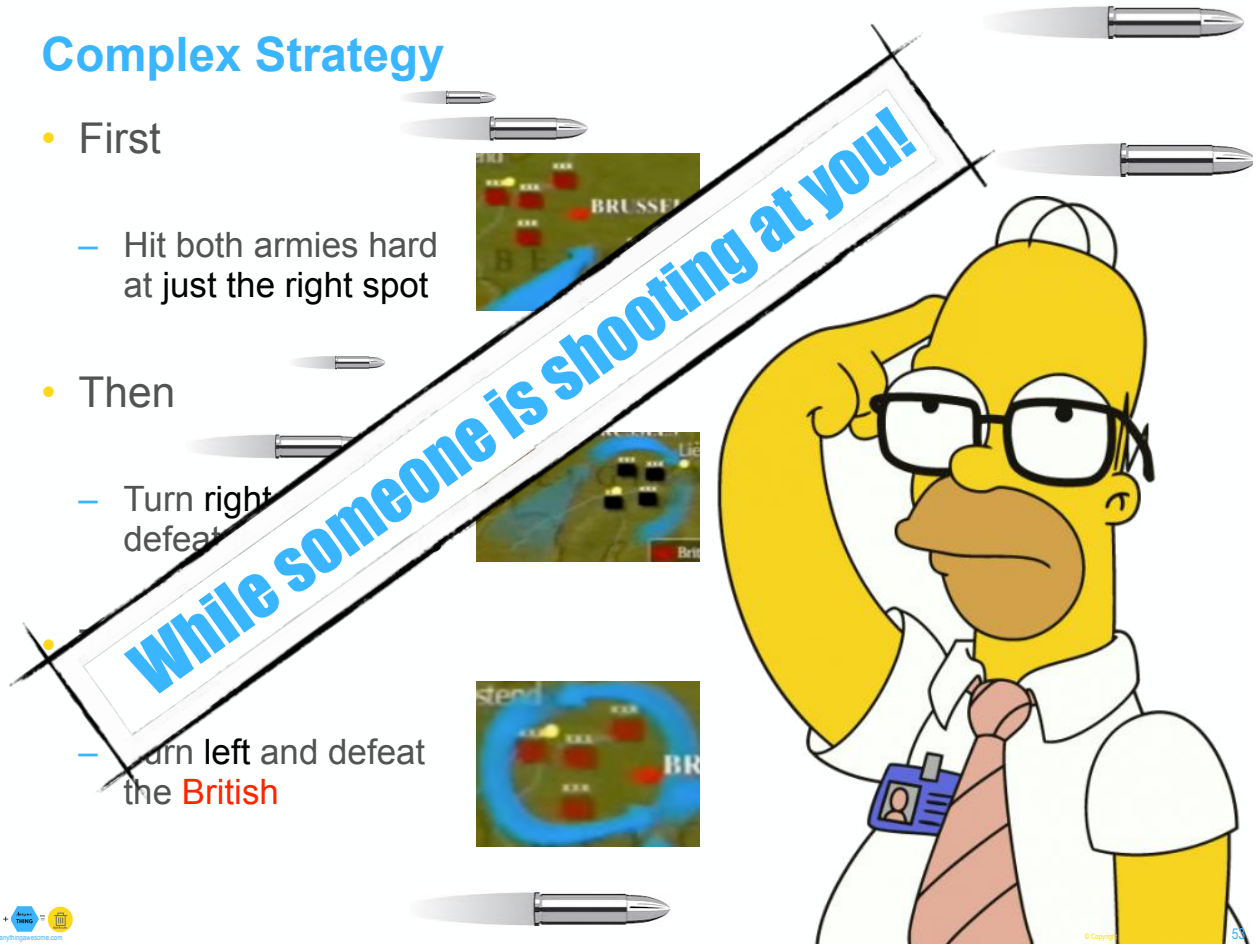
Every Day Low Price

Strategy in Action: Napoleon Faces a Larger Enemy

- Question?
  - How do I defeat the competition when their forces are bigger than mine?
- Answer:
  - Divide and conquer!
  - **"a pattern in a stream of decisions"**

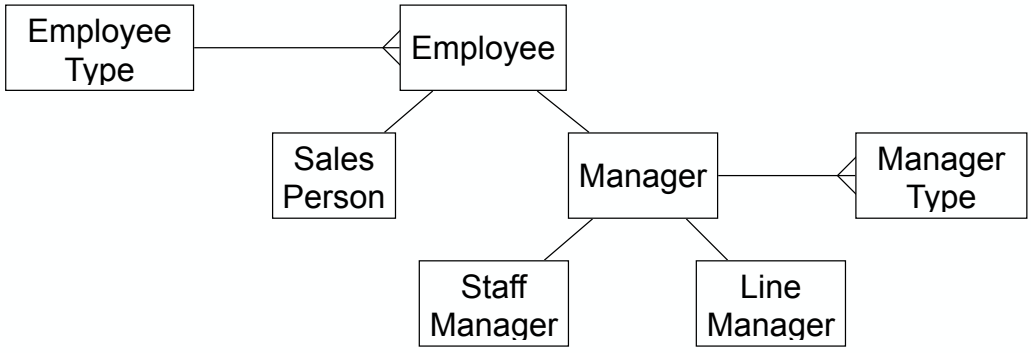
Complex Strategy

- First
  - Hit both armies hard at just the right spot
- Then
  - Turn right and defeat
- Then
  - Turn left and defeat the British



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

Strategic Use of Data Models (Examples)

- SABRE creates flight booking business

Sabre

An innovation technology company

AT&T

- AT&T invents the "new" credit card business overnight

- Amazon invents at home retailing

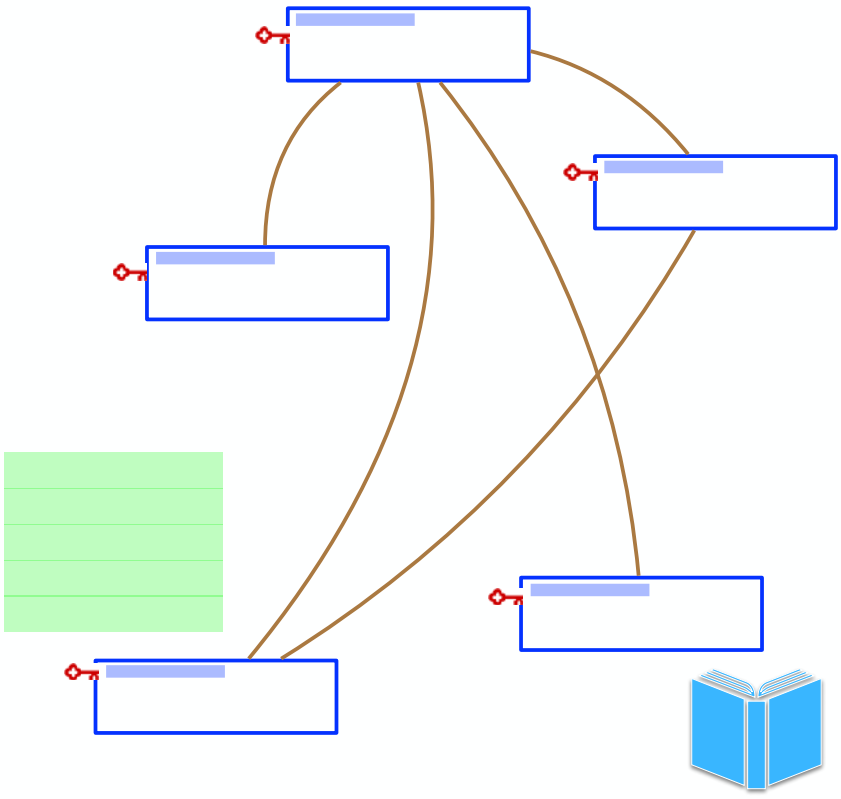
amazon.com

- CapitalOne reinvents solicitation

Capital One | what's in your wallet?

Data Modeling Process

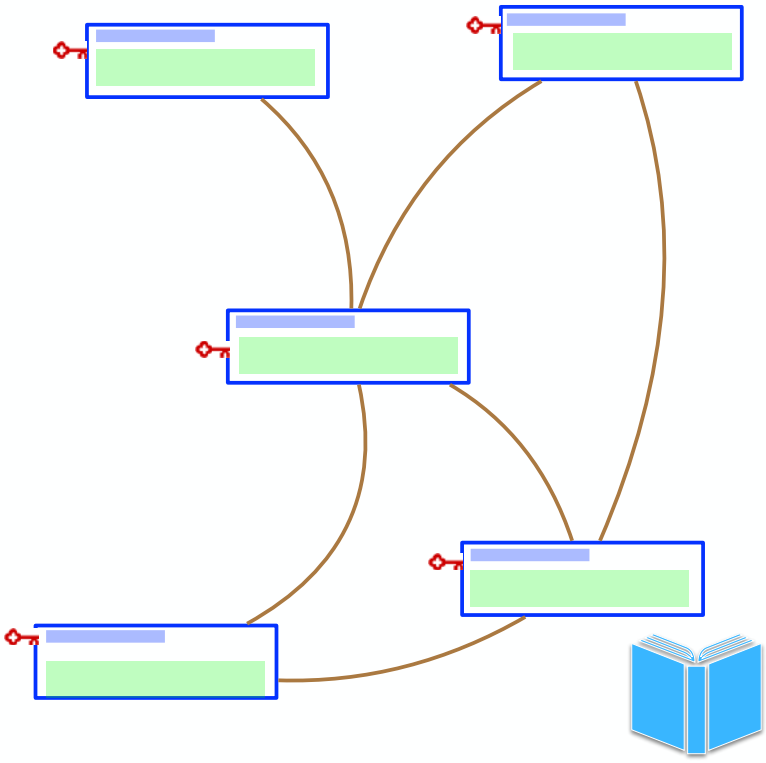
1. Identify entities
2. Identify key for each entity
3. Draw rough draft of entity relationship data model
4. Identify data attributes
5. Map data attributes to entities



Trusted Catalog

Model Evolution Is Good, at First ...

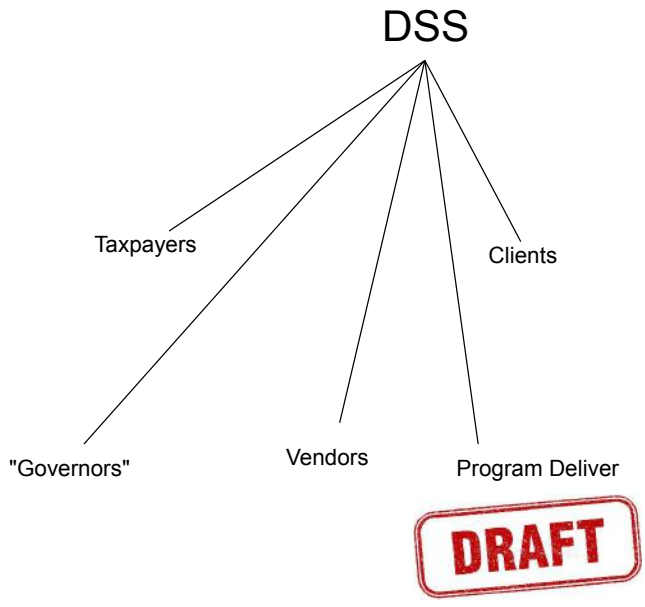
1. Identify entities
2. Identify key for each entity
3. Draw rough draft of entity relationship data model
4. Identify data attributes
5. Map data attributes to entities



This Logical Data Model Is Comprised of 5-Model Views

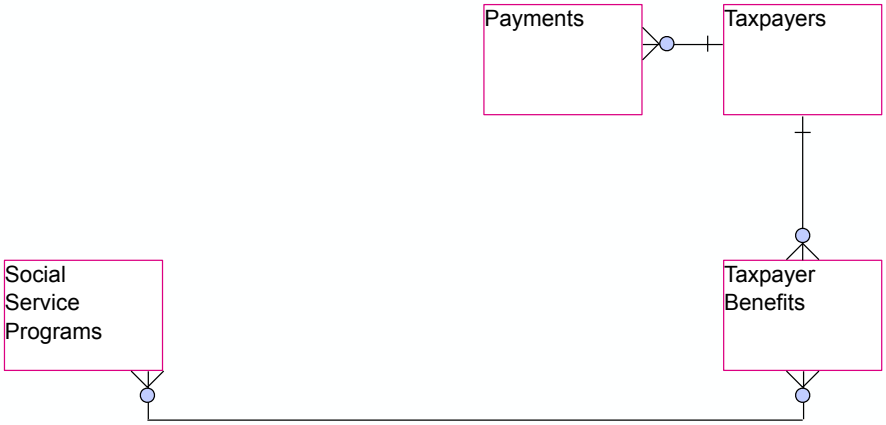
DSS Strategic Data Model

- Taxpayer view
- Client view
- Governance view
- Program Delivery view
- Vendor view

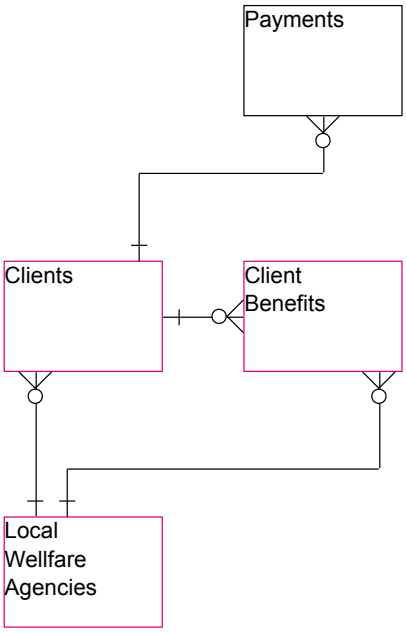


(Please note that all models are currently unvalidated and should be consider as "draft" version until they are validated!)

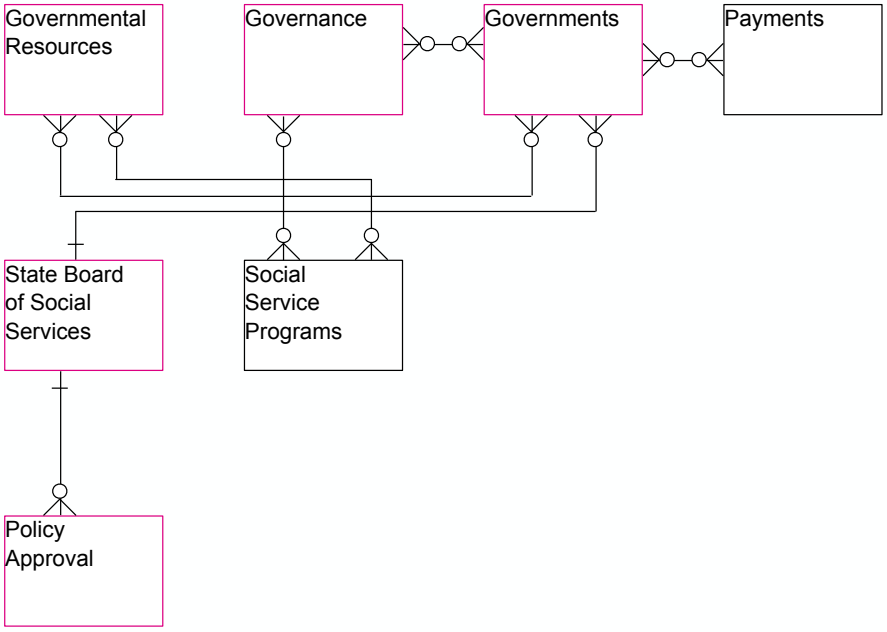
Taxpayer View



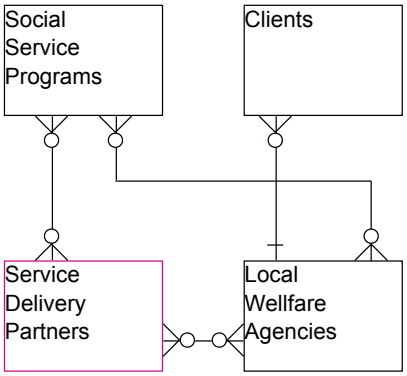
Client View



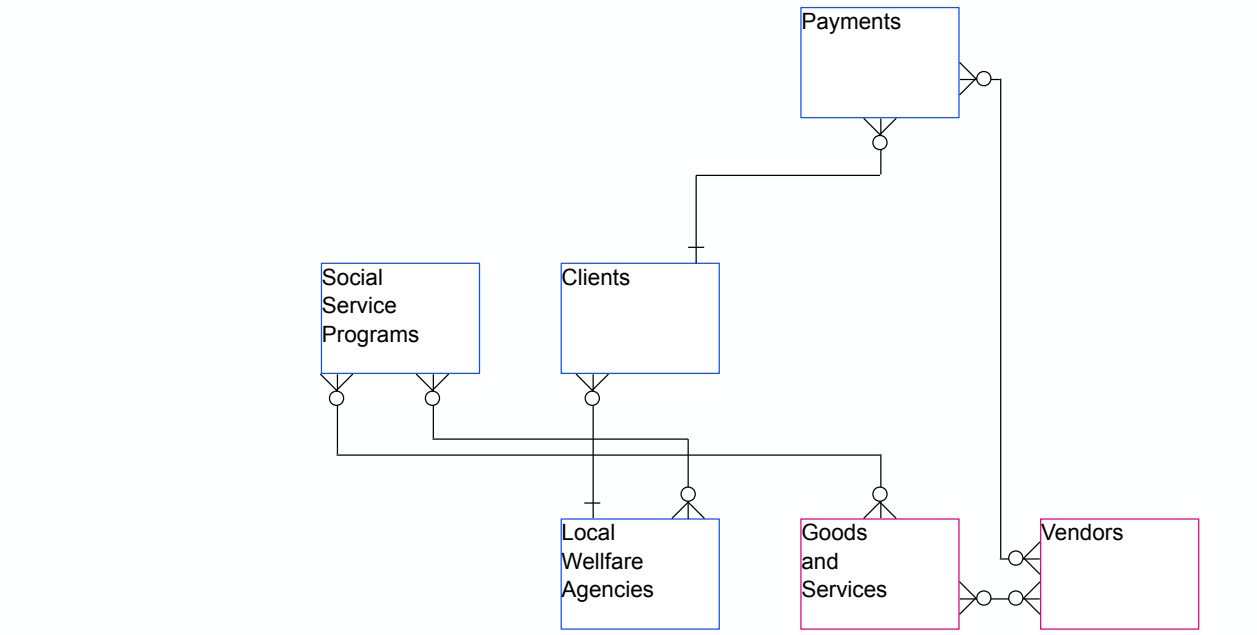
Governance View



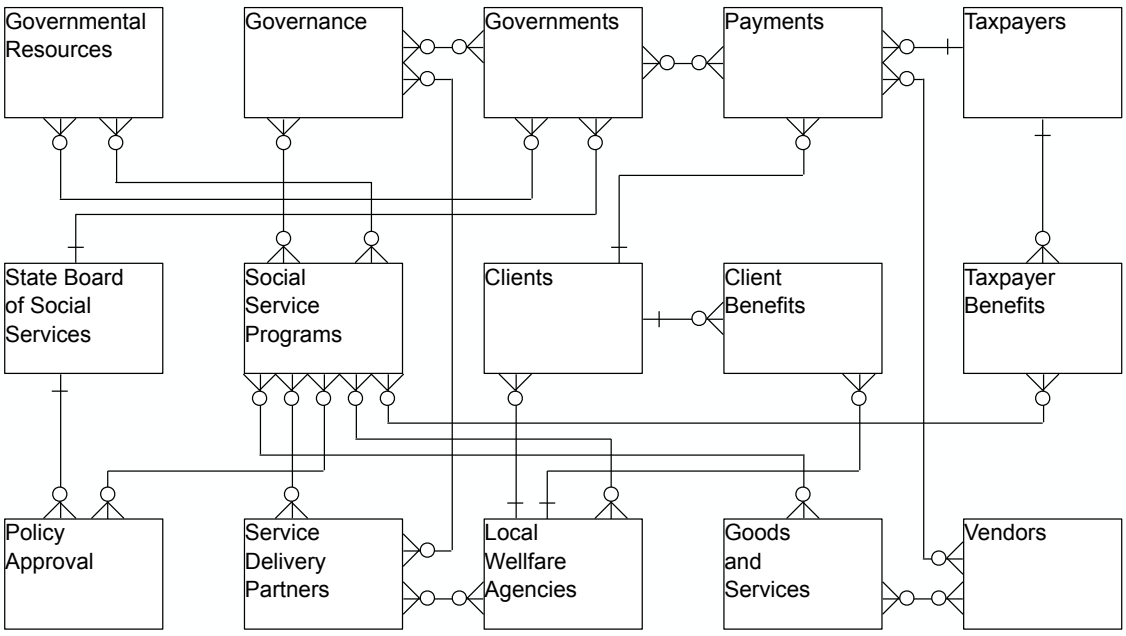
Program Delivery View



Vendor View



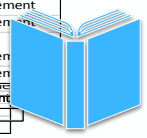
DSS Conceptual Data Model



Business Glossary

- Start of Enterprise Taxonomy
- Defines Initial Entities for Conceptual Data Model
- Engages the Business Community to Validate Entities and provide meaningful business definitions

Entity	Description	Domain Area
Donor	Funder	Business Development
Solicitations	Need for Work	Business Development
Solicitations Proposal	Response to Need for Work	Business Development
Pre-Positioning	Intelligence Gathering	Business Development
Award/Sub-Award	Funding Vehicle	Business Development
Terms Conditions	Details about a Funding Vehicle	Business Development
Budget	Amount of Money Available	Business Development
Work Plan	Set of Activities to Complete	Business Development
PMP	Monitoring Plan for Activities	Business Development
Project	An NGO Project is defined as a self-contained set of interventions or activities with the following characteristics: a) an external client; b) purchase order, contract or agreement; c) expected deliverables, outcomes and results; d) a beginning and end date of implementation; e) an approved budget; and full and/or part time NGO staff	Project Management
Geographic Area		Project Management
Office Locations	Location in which a Central Office resides	Project Management
Project Roles		Project Management
Project Artifacts		Project Management
Project Budget		Project Management
Project Work Plan		Project Management
Milestones	Schedule of completed activities	Project Management
Monitoring	Plan to measure Activities	Project Management
Evaluation	Assessment of Activities	Project Management
Indicators	Target of Outcome	Project Management
Outcomes	Statement of what needs to be accomplished	Project Management
Acct Receivable	Payments to NGO	Financial Management
Chart of Accounts	Defined Accounts	Financial Management
Payroll	Process to Pay Worker	Financial Management
Supplier	Provider of Goods or Service	Financial Management
Contract	Binding Agreement	Financial Management
Purchase Order	Statement of Good or Service	Financial Management
Performance	Level of Success	Talent Management
Benefits		Talent Management
Skills		Talent Management
Worker	Person who has been hired by NGO	Talent Management
Candidate	Potential hire of NGO	Talent Management

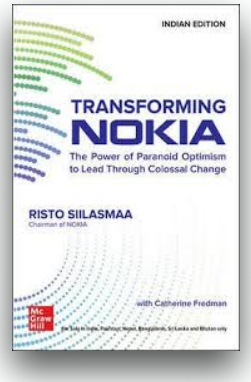


Trusted Catalog

(Pre Microsoft Acquisition)

NOKIA

- Tires, rubber products
- Consumer electronics
- Mobile phones
  - Finns are bilingual (2% of population speaks Swedish)
  - Nokia wanted to play internationally
  - English mandated in all business settings
  - Lots of words were unknown
  - Culturally: Bad to not ask questions
  - Culturally: Good to build common vocabulary
- When an unfamiliar term was used
  - Group: Access NTB to see if there existed a golden definition
  - Group: If not, vote whether to submit it for inclusion in the NTB
  - Weekly: the NTB group reviewed submissions
  - Weekly: the NTB group published new versions of the NTB
  - NTB = Nokia Term Bank



NTB = Trusted Catalog

Cruiser Collector

NOKIA

NOKIA CRUISER-COLLECTOR IN CAPITAL AREA

The Cruiser located by your desk for sorting waste has three sections:

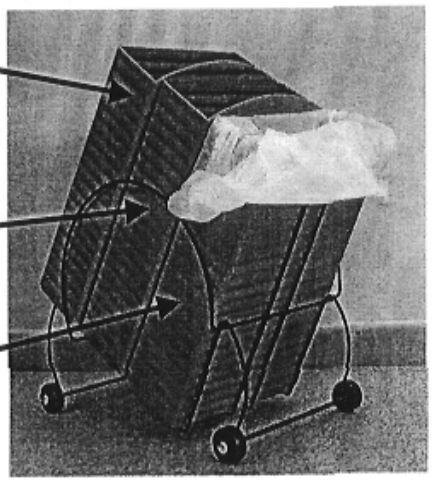
1. Office paper for shredding
- all white office paper, also the printer cover pages
- all white-based copy-paper
- all white-based printing paper
- all white memo slips

Take your confidential papers to the locked container in the office service point.

2. Recycling paper (newspaper, journals)
- newspaper and journals
- advertisement
- non-confidential coloured paper
- envelopes

Take your recycling papers to the non-locked container in the office service point.

3. Mixed waste
- rubbish
- plastic folders
- stickers
- Post-it slips
- wrap around the office paper reams



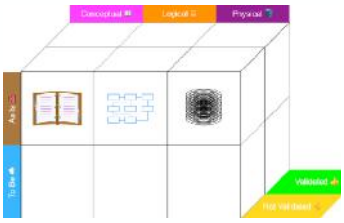
Sort your papers right for information security and environment!

Take your biowaste to the container in the floor

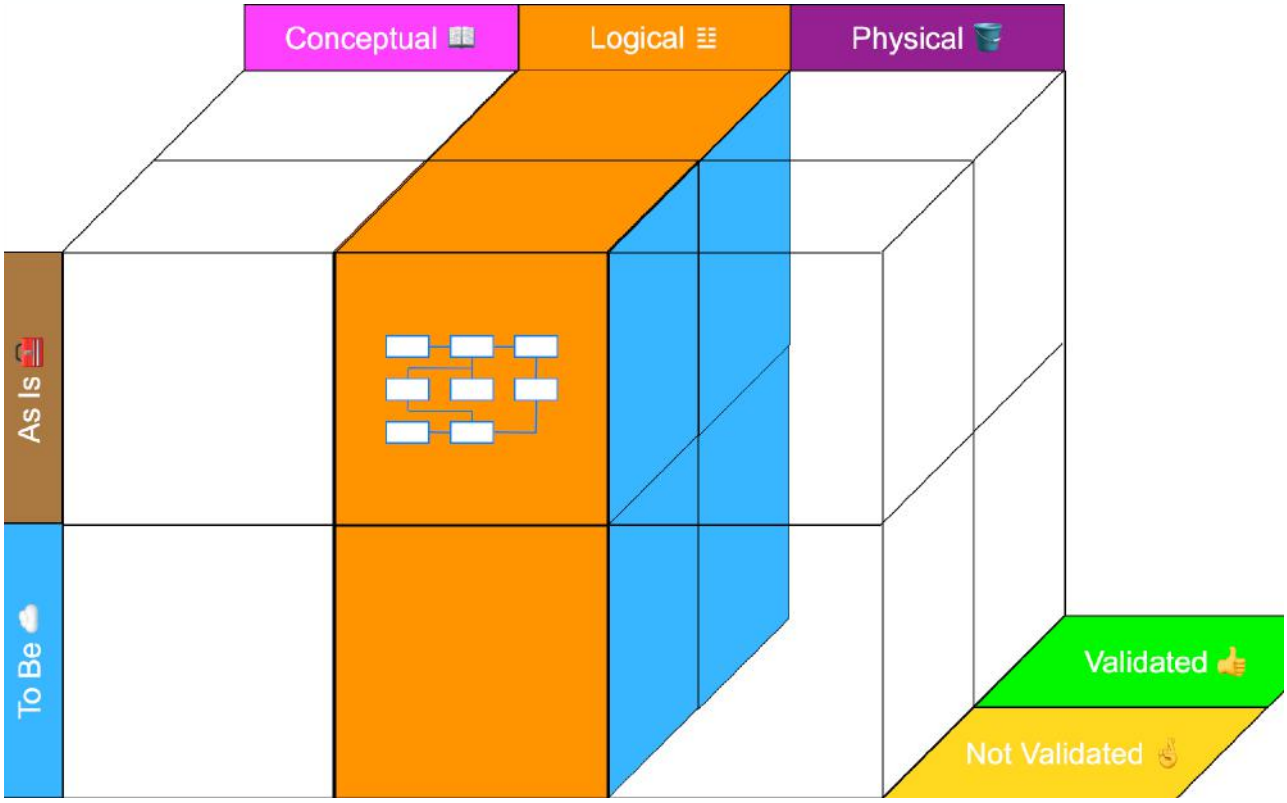
Program Overview

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

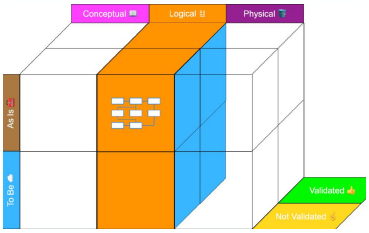
- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A



Logical Data Modeling



Logical Data Modeling



Motivation

- Provide data specification information about effort
  - Size
  - Shape
  - Provenance
  - Functions
  - Down stream uses
- Free discussions from technological considerations that are separate from business objectives
- Document preliminary data designs satisfying business objectives
- Generate as much as possible

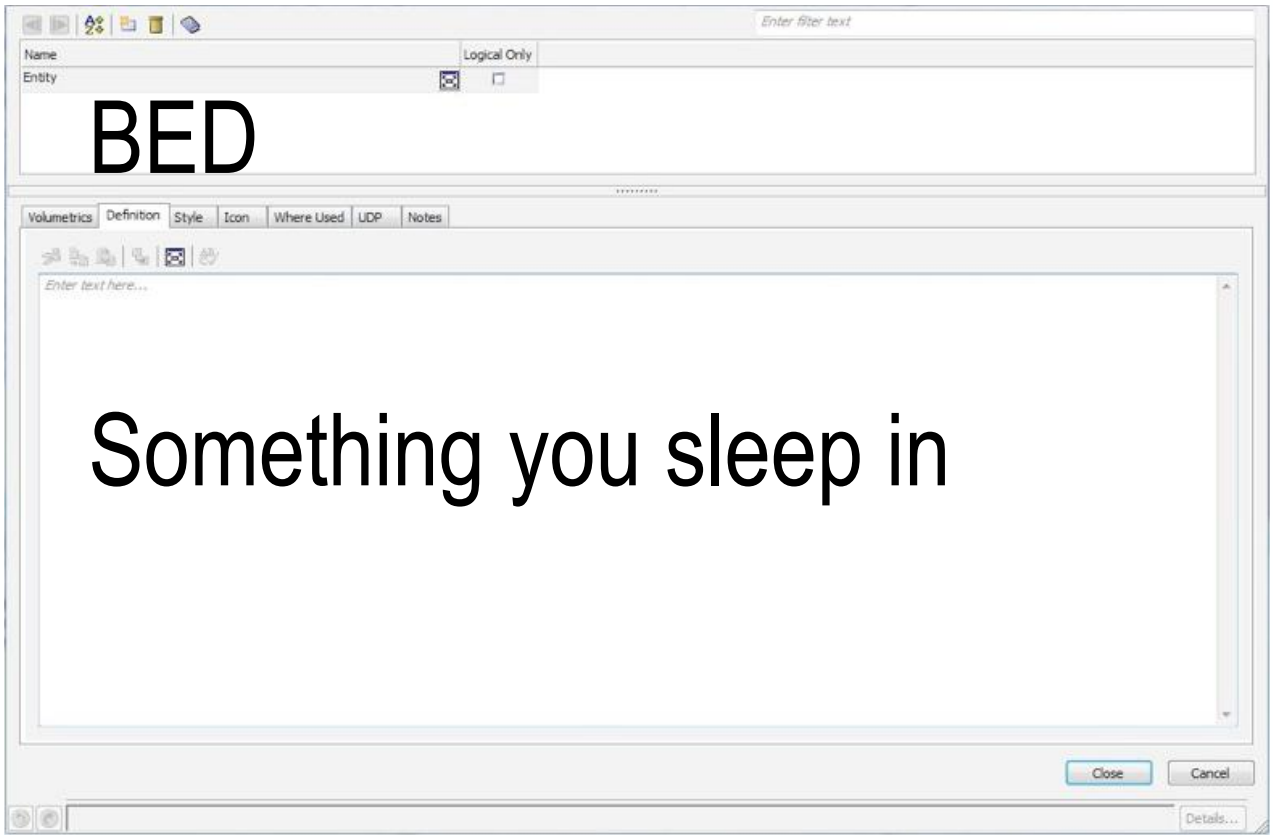
As Is Logical Data Models

- Challenge the conceptual model (if it exists)
- Explicitly incorporate relevant information from existing components

To Be Logical Data Models

- Serve as the organizing principle around which system data capabilities are built
- Facilitates common vocabulary among business and technical analysts

Standard Definition Reporting Does Not Provide Conceptual Context



Purpose Statement Incorporates Motivations

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

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; and
- Associations with other data items(read as "One room contains zero or many beds.")

Q: What Is the Proper Relationship for These Entities?

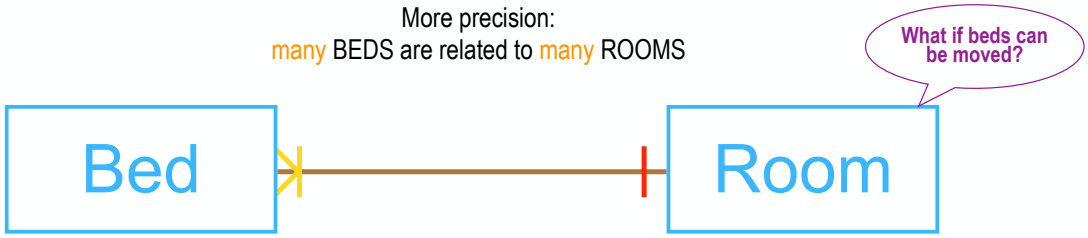
Bed

Room

Data Maps at the Entity Level ➜ Stored Facts

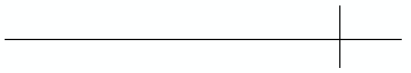
a BED is related to a ROOM

More precision:
many BEDS are related to many ROOMS

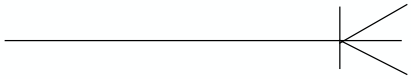


Better information:
many BEDS may be contained in each ROOM and each room may contain many beds

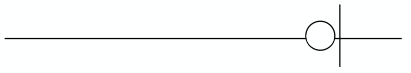
Possible Entity Relationship Cardinality Options



Exactly One (mandatory)

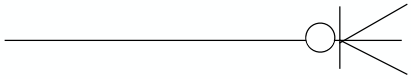


One or Many (mandatory)



Eventually One (optional)

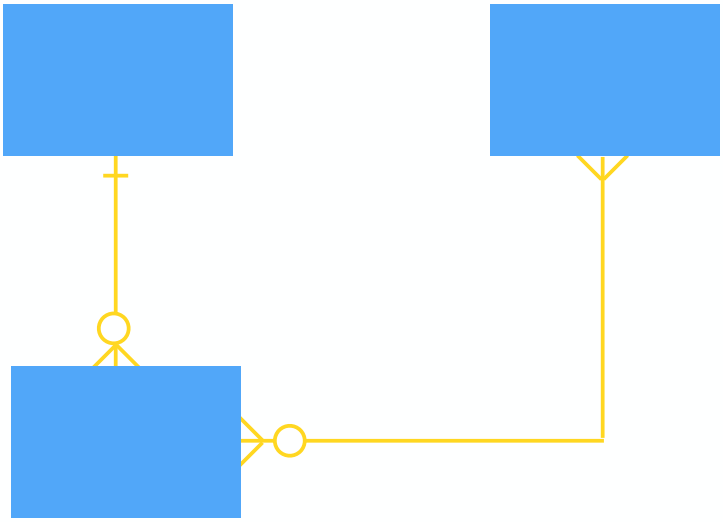
Zero, or Many (optional)



Eventually One or Many (optional)

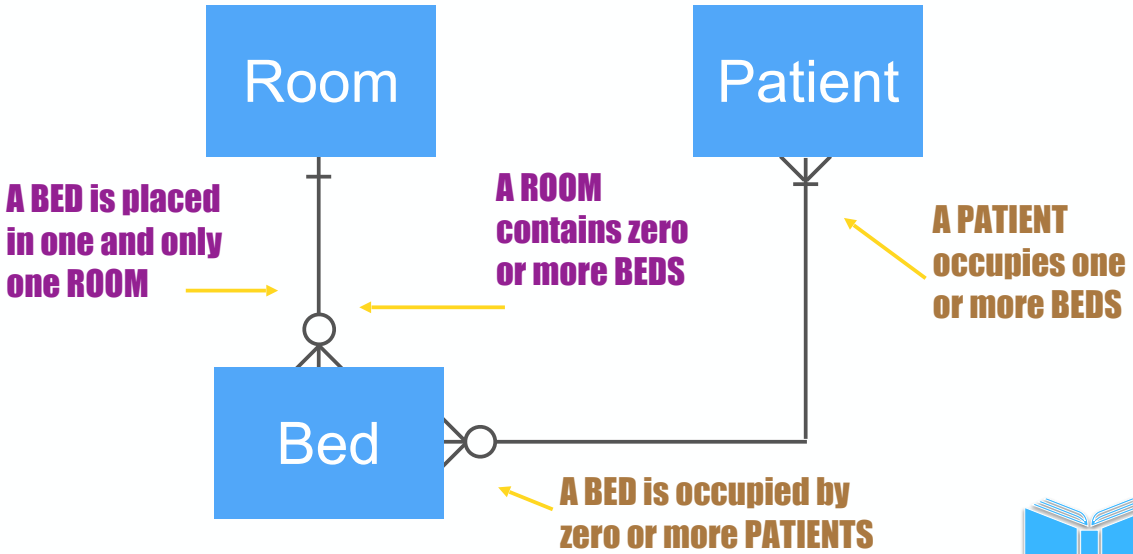
What Is a Relationship?

- Natural associations between two or more entities



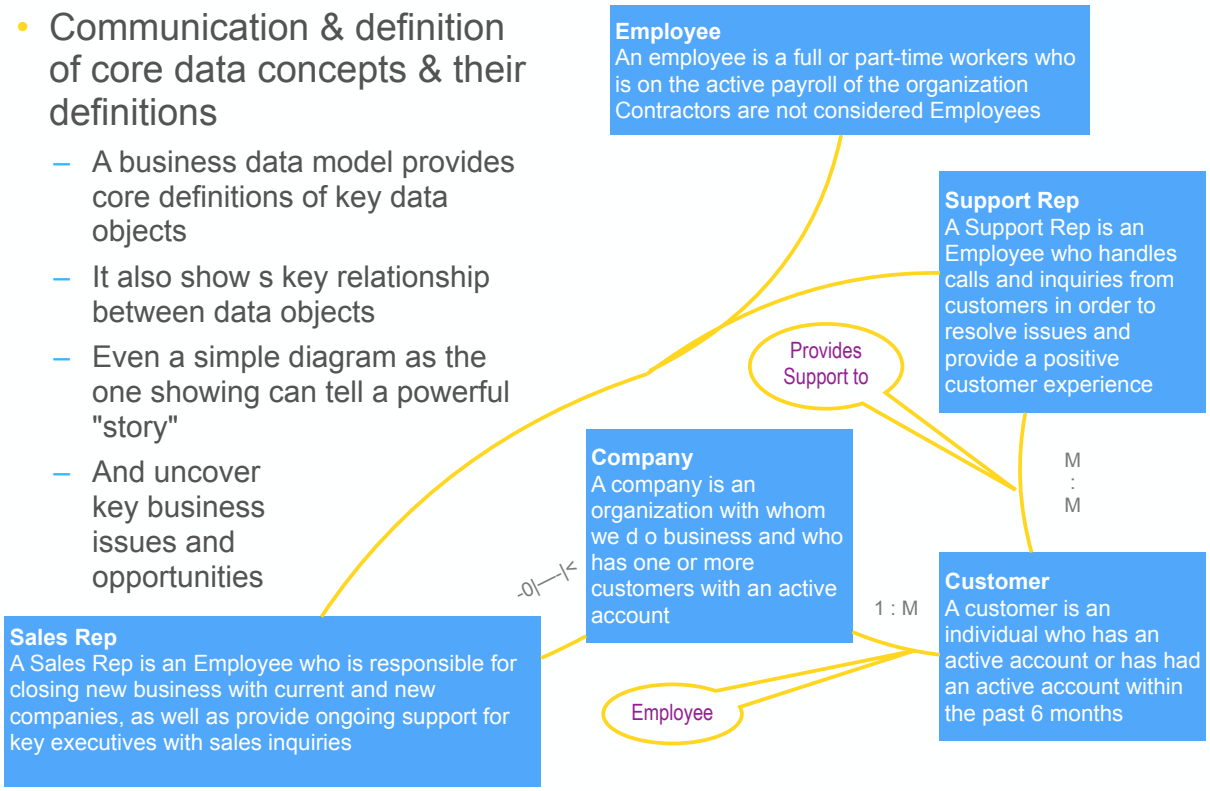
Ordinality & Cardinality (Refinements on Relationships)

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

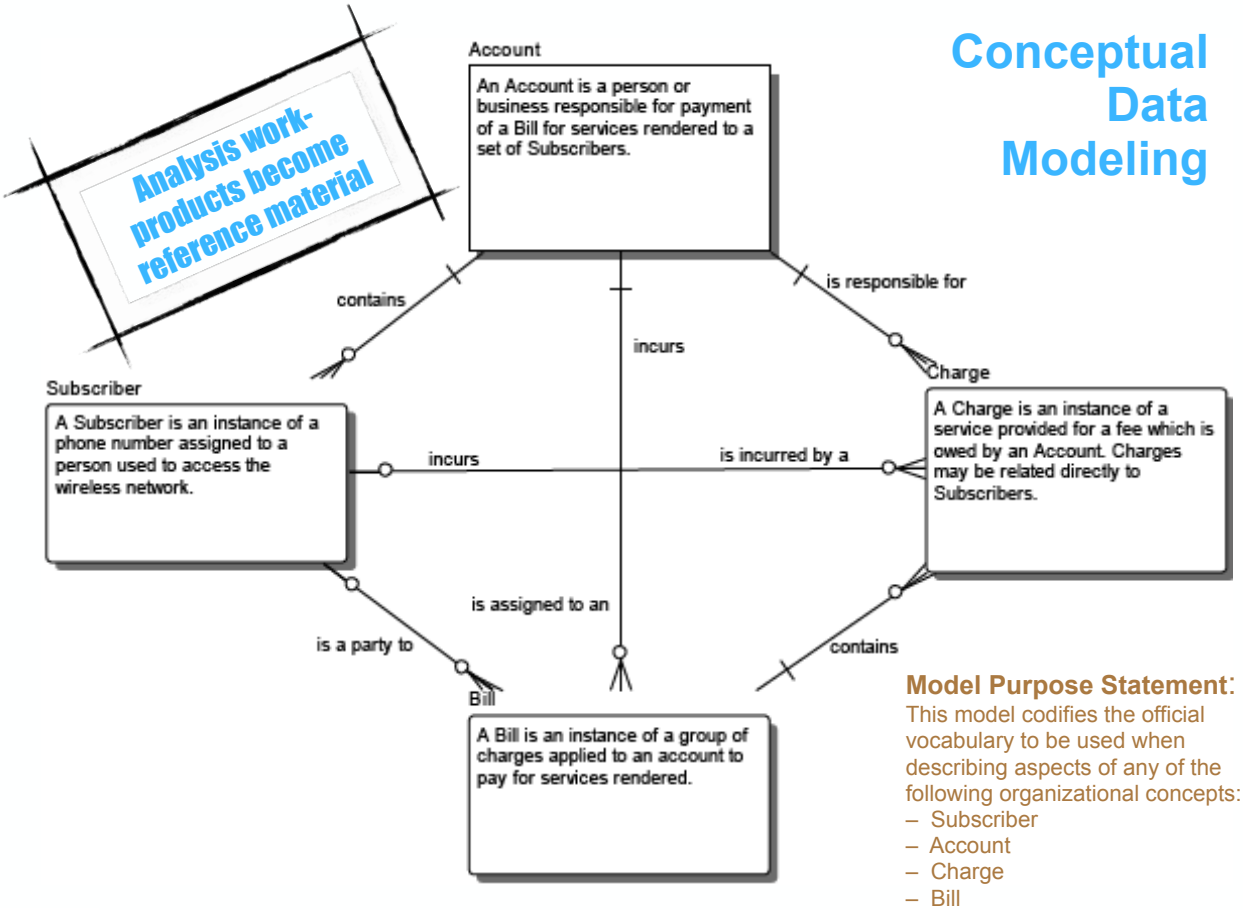


Business Data Model (Conceptual)

- Communication & definition of core data concepts & their definitions
  - A business data model provides core definitions of key data objects
  - It also show s key relationship between data objects
  - Even a simple diagram as the one showing can tell a powerful "story"
  - And uncover key business issues and opportunities



Conceptual Data Modeling



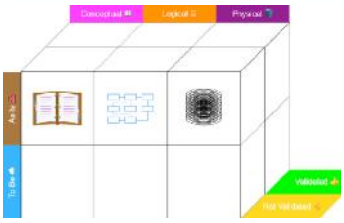
Model Purpose Statement:
This model codifies the official vocabulary to be used when describing aspects of any of the following organizational concepts:
- Subscriber
- Account
- Charge
- Bill

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

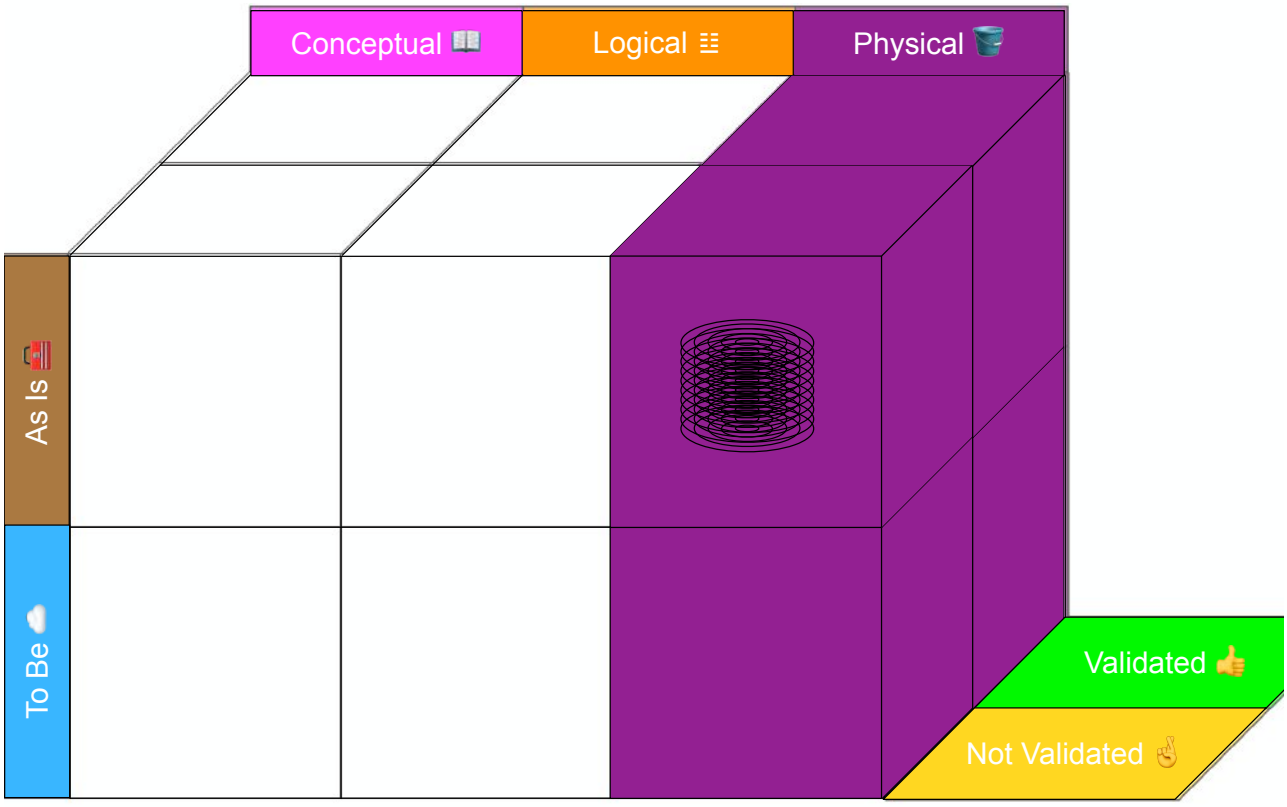
Program Overview

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

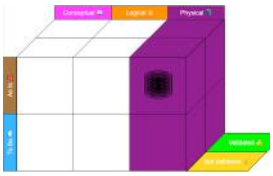
- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A



Physical Data Modeling



Physical Data Modeling



Motivation

- Documentation of specifications of production systems
  - Data flow diagrams
  - Entity-relationship diagrams
  - Dictionary/Glossary/Catalog
- Should exist if system is in production
  - Why would anyone handmake DDL with today's tool capabilities?
- Must exist to create the system that is put into production
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution

As Is Physical Data Models (Exist too)

- This should be foundational system documentation
- Description required to access data 'in the system'
- Often can be reverse engineered, semi-automatically

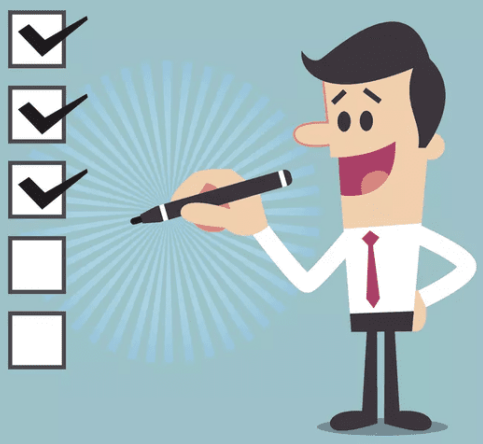
To Be Physical Data Models (Exist too)

- This is a specification of the data that can be accessed by the application
- Specification of current and future data elements to be maintained by application
- Often can be generated, semi-automatically

How Is Data Stored and Represented?

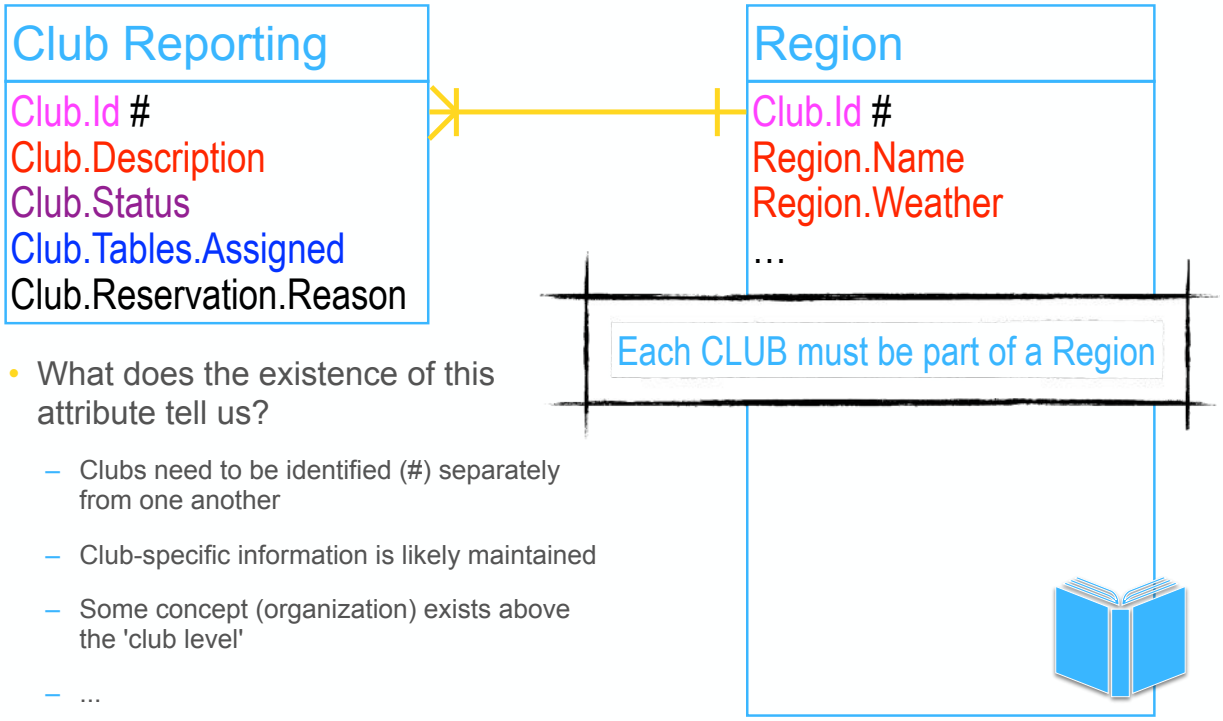
- Lists of organizational **persons places things** that need to be **created read updated deleted archived**

- These are called Attributes
  - Attributes are characteristics of "things"



Analyzing Data Attributes and Relationships

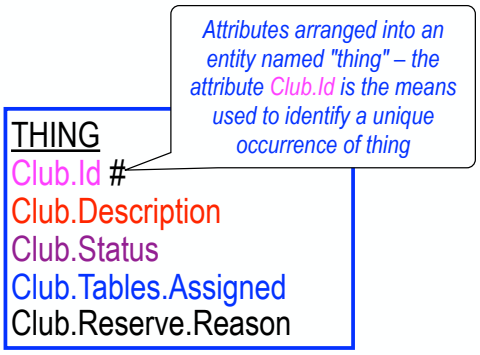
- Characteristics of CLUBS and REGIONS

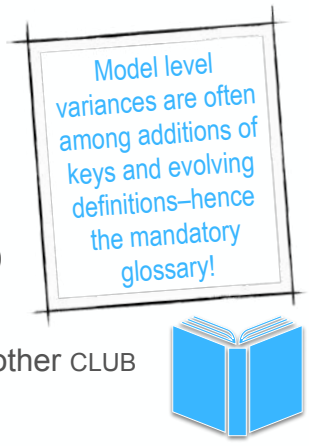


- What does the existence of this attribute tell us?
  - Clubs need to be identified (#) separately from one another
  - Club-specific information is likely maintained
  - Some concept (organization) exists above the 'club level'
  - ...

Data Modeling Uses

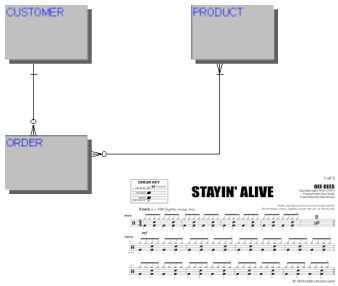
- An organization might decide to characterize the parts of a THING as:
  - Attributes: ID, description, status, Tables.Assigned, reserve.reason
- Decisions to manage information about each specific attribute has direct consequences
  - A decision to use the above data attributes permits the organization to determine if it has tables are available to be reserved
- Characteristics can be shared
  - All CLUBS may have a status
  - Many REASONS can be assigned to reservation (free text)
- Characteristics may be required to be unique
  - ID permits identification every CLUB as distinct for every other CLUB
  - Description is likely to be unique for each CLUB

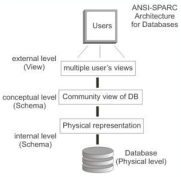




Data Modeling Requirements

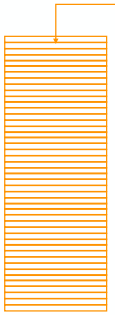
- The process of discovering, analyzing, and scoping data requirements
  - Understand what the data things are?
  - What do they do?
  - How do they interact?
- Representing/communicating requirements in a precise form called a data model
  - Maps of critical business assets
  - Compose and contain metadata essential to data consumers
  - Function as a kind of sheet music language
  - Metadata is essential to other business functions (definitions for governance, lineage for analytics, etc.)
- The process is iterative and may include conceptual, logical, and physical models
- Modeling is done to accomplish a goal!





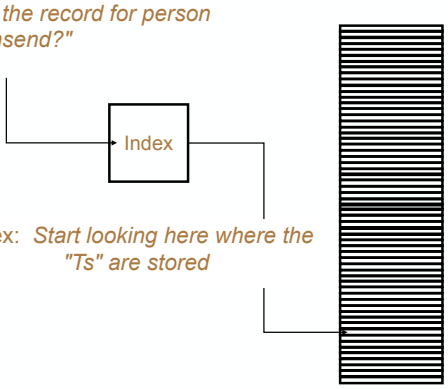
5 Basic Database Structures

Program: *Must start at the beginning and read each record when looking for person "Townsend?"*



Flat File: Records are typically sorted according to some criteria and must be searched from the beginning for each access

Program: *Where is the record for person "Townsend?"*

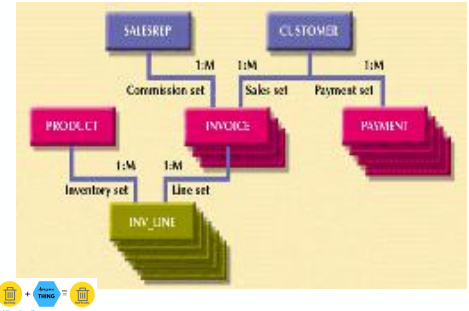


Index: *Start looking here where the "Ts" are stored*

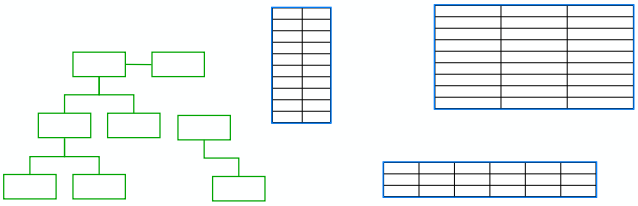
Indexed Sequential File: Built-in index permits location of records of persons with last names starting with "T"

Associative Concept-oriented, Multi-dimensional, XML database, 3NF, Star schema, Data Vault, graph, LakeHouse

Network Database: Records are related to each other using arranged master records associated with multiple detail records using linked lists and pointers

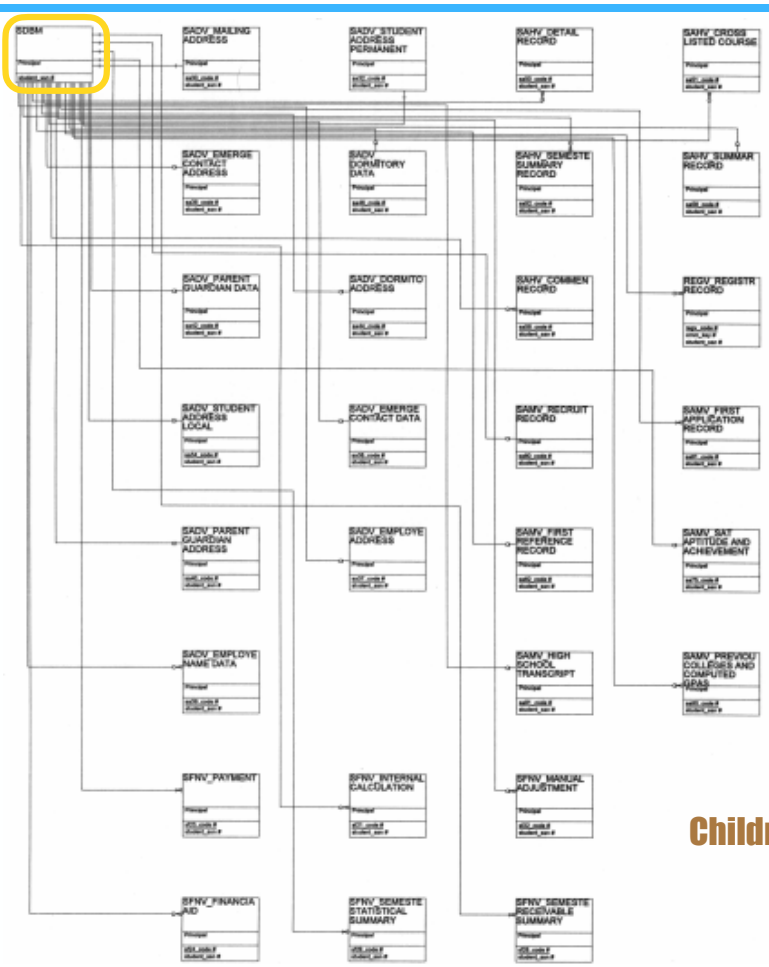


Relational Database: Records are related to each other using relationships describable using relational algebra



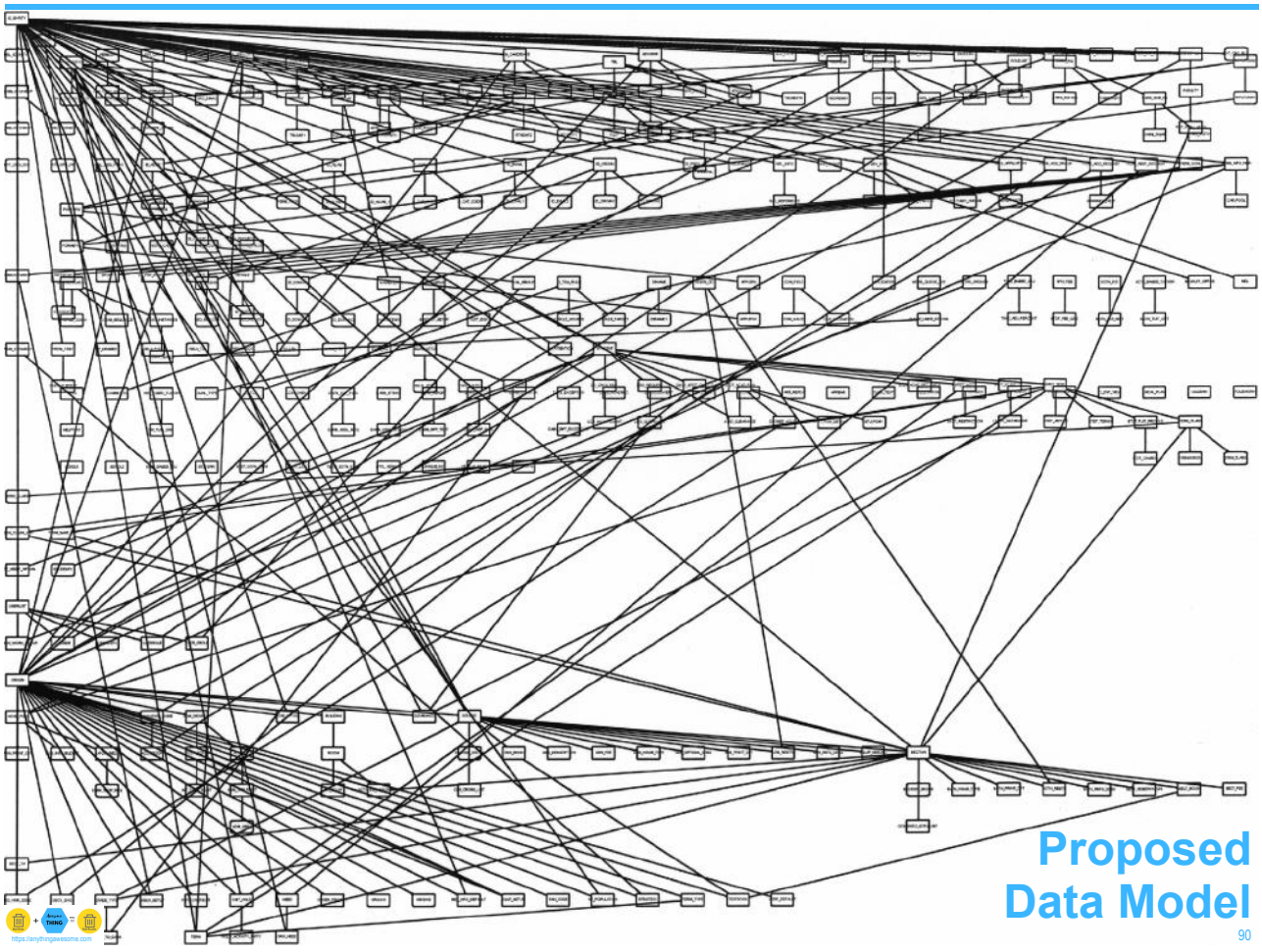
Hierarchical Database: Records are related to each other hierarchically using 'parent child' relationships

Parent



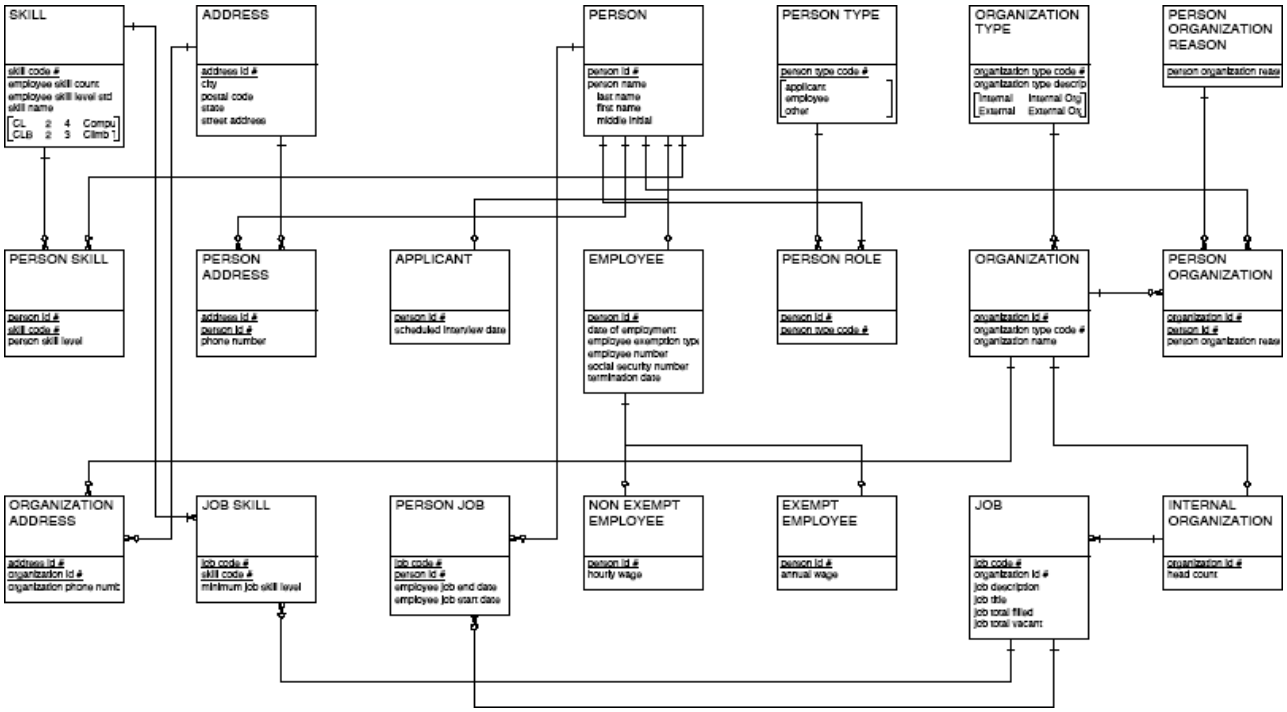
Student Data Base Master

Children

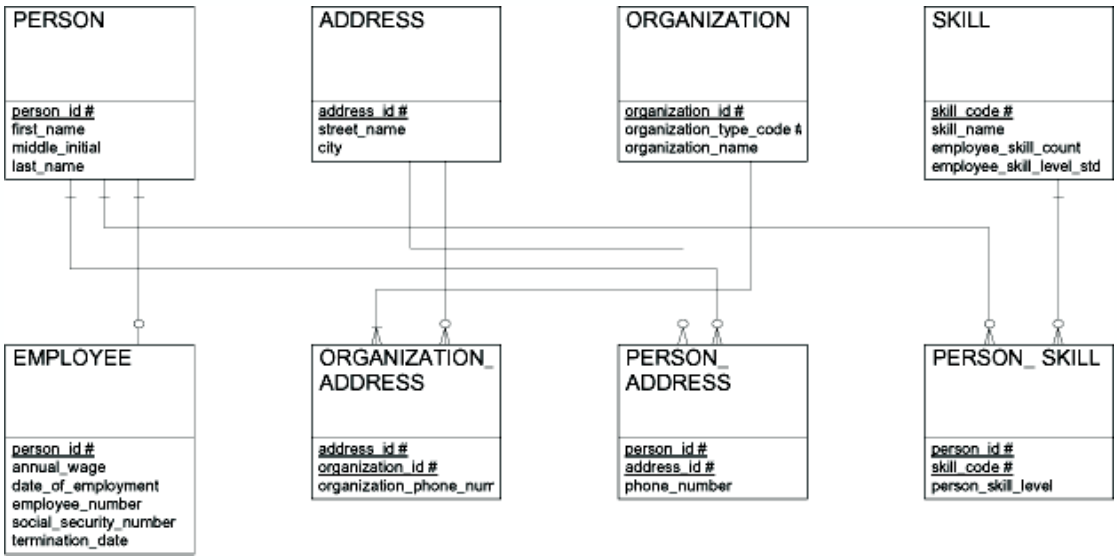


Proposed Data Model

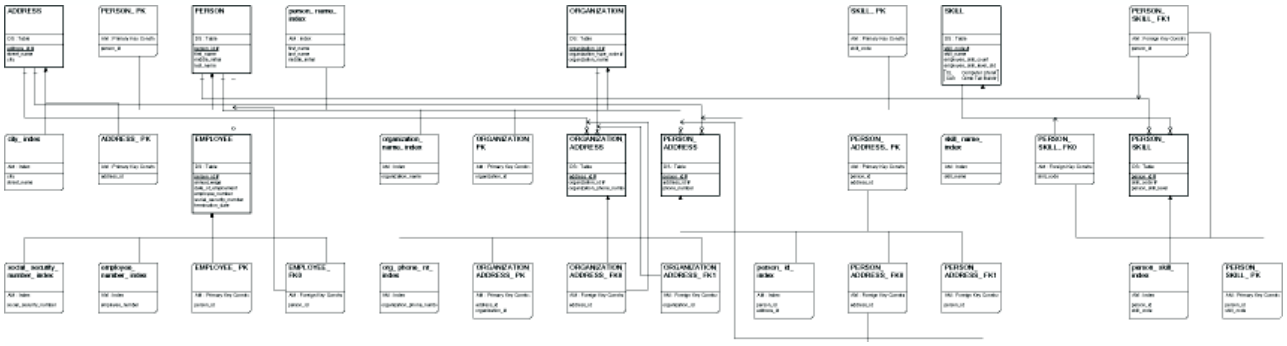
HR Conceptual Model

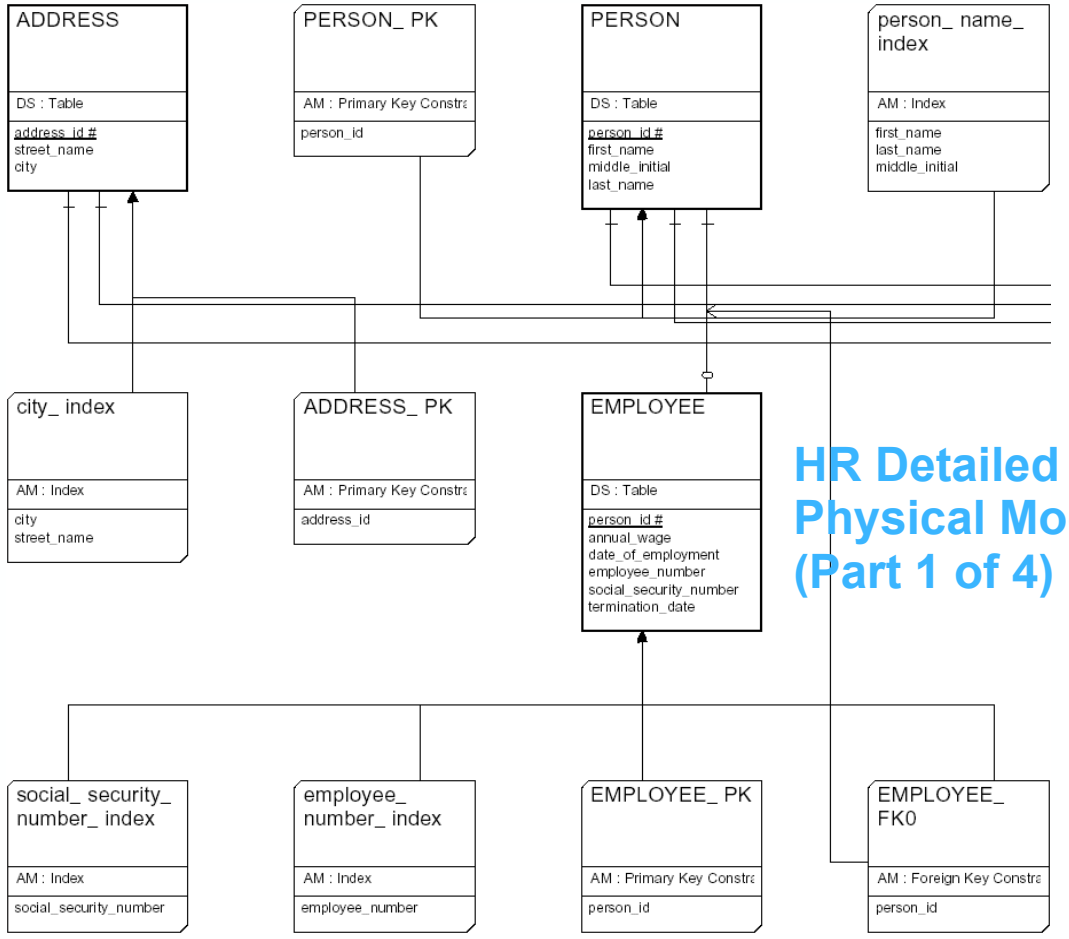


HR Logical Model



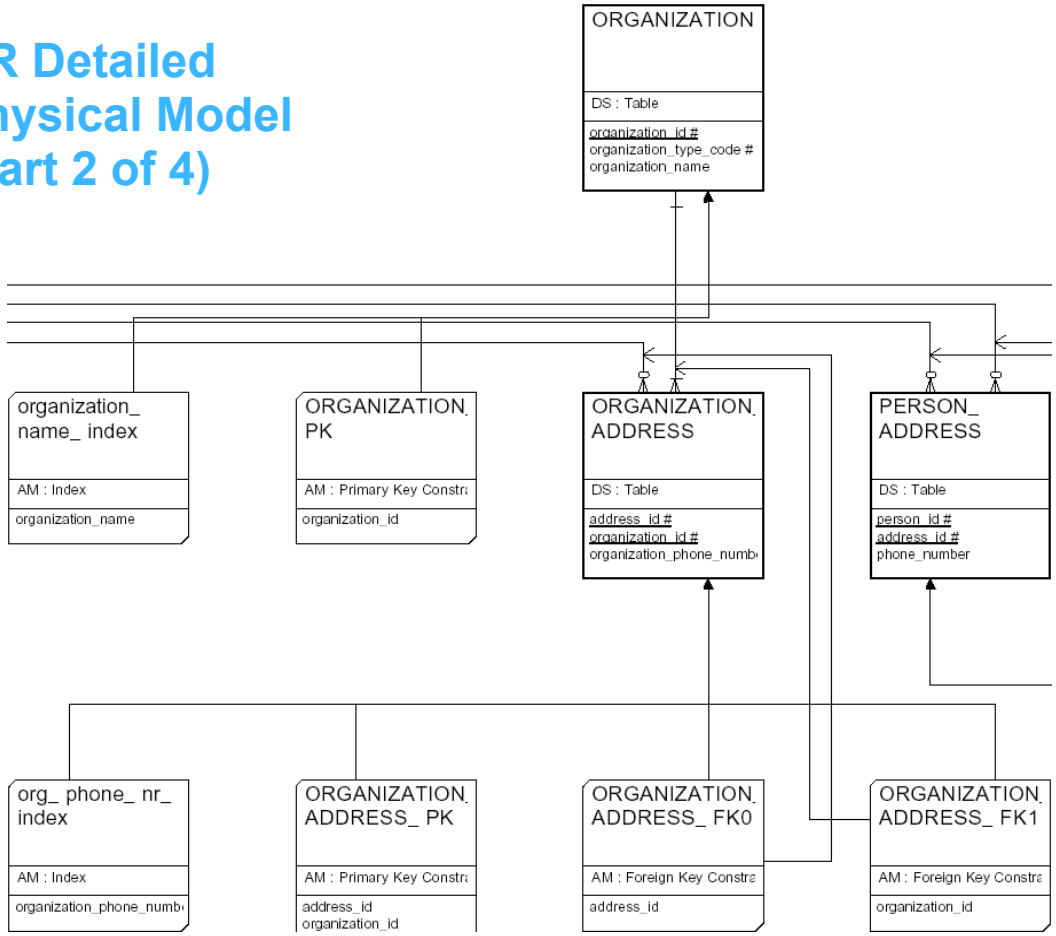
HR Detailed Physical Model Overview

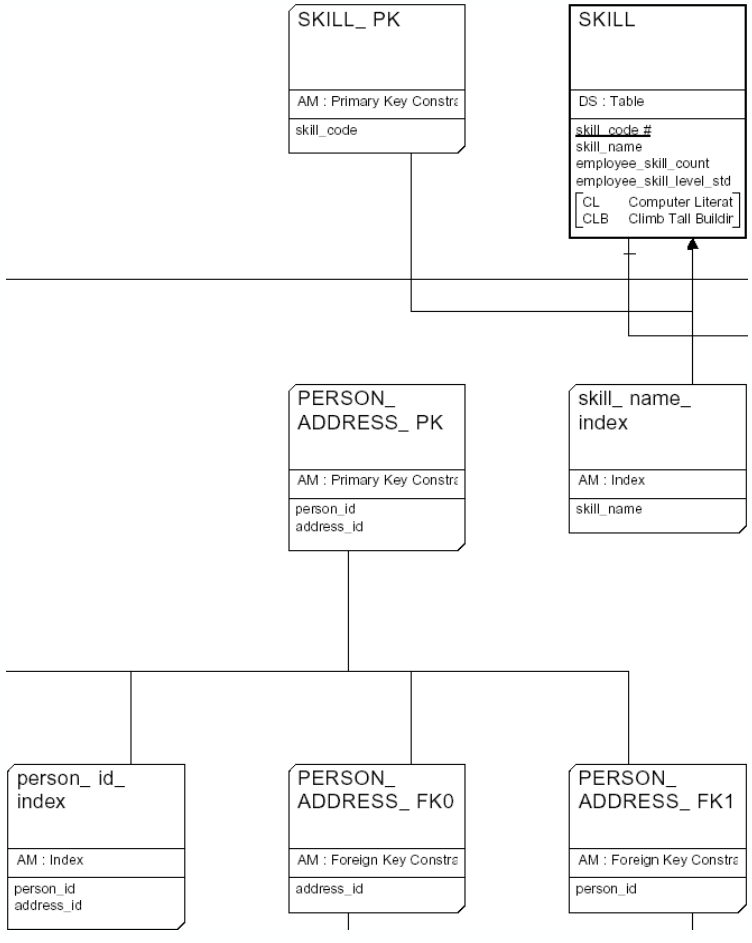




HR Detailed Physical Model (Part 1 of 4)

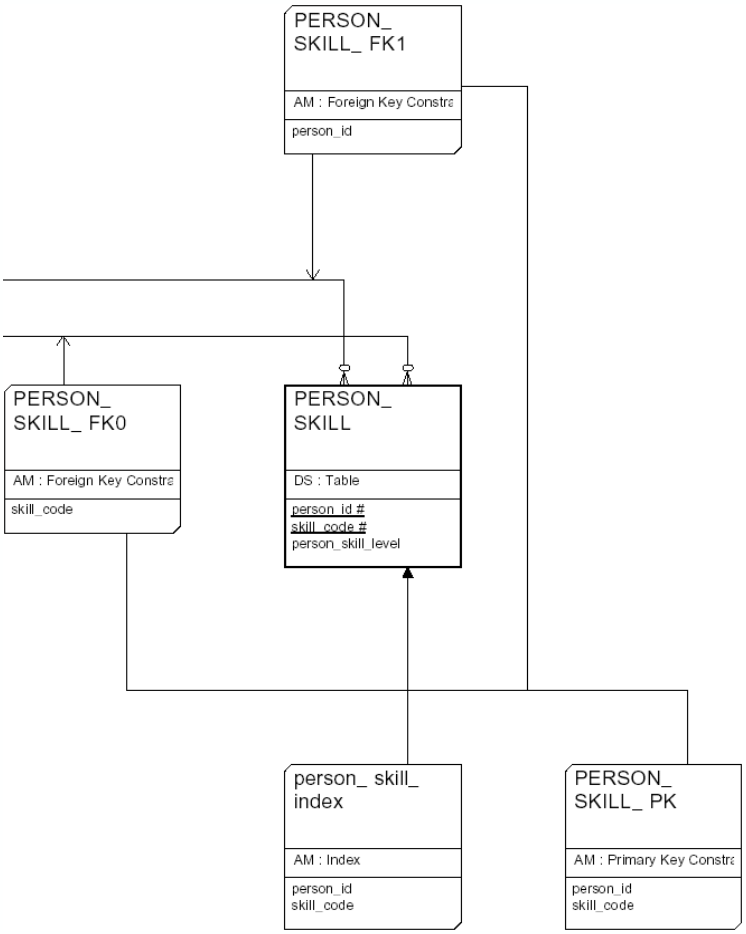
HR Detailed Physical Model (Part 2 of 4)





HR Detailed Physical Model (Part 3 of 4)

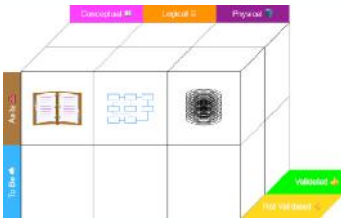
HR Detailed Physical Model (Part 4 of 4)

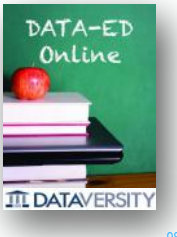


Program Overview

Conceptual
Versus
Logical
Versus
Physical
Data Modeling

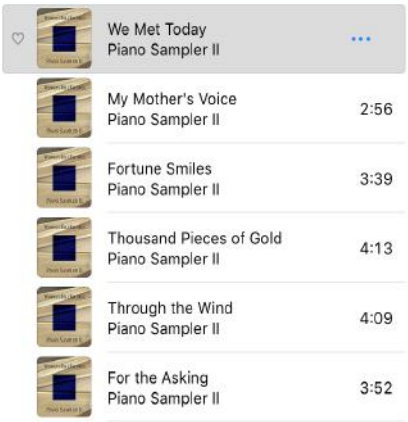
- Introduction to Modeling Data
  - Motivation
  - 3 primary data model types (+ plus two characteristics)
  - Reasons for each
  - Purposeful Modeling Basics (conversions, forward/reverse engineering)
- Conceptual
  - Motivation: Architectural tradeoffs
  - Strategy and conceptual data modeling
  - Glossary/Dictionary capabilities
- Logical
  - Motivation: Simplicity (Operational and Design)
  - Motivation towards standards
  - Business meets strategy
- Physical
  - Motivation: Required documentation and/or facts
  - Become the blueprints for physical construction of the solution
  - Blueprints are used for future maintenance of the solution
- Take Aways/References/Q&A





There Are Correct Ways To Organize Data

- All involve data modeling
- Optimization can be done for:
  - Flexibility
  - Adaptability
  - Retrievability
  - Risk reduction
  - ...
- Techniques include:
  - Data integrity
  - Smart codes bad/dumb codes good
  - Architecture (table joins)
  - ...



Don't Tell Them That You Are Modeling!

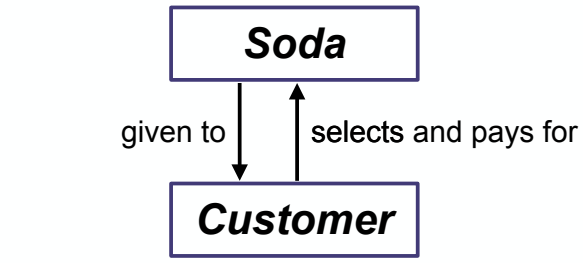
Just write some stuff down

Then arrange it

Then make some appropriate connections between your objects

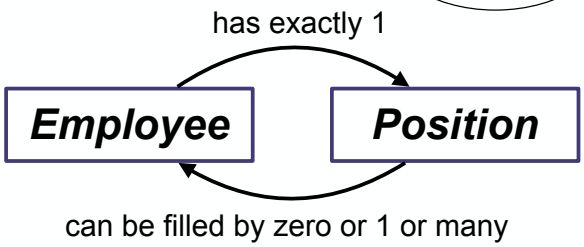
Keep Focused on the Data Model's Purpose

- The reason we are locked in this room is to:
  - Mission: *Understand formal relationship between soda and customer*
    - Outcome: Walk out the door with an as is physical and logical 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 characteristics that represent the brand with a logical data model
  - 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 - need conceptual model for board presentation



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*

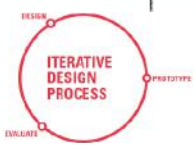


Inspired by: Karen Lopez http://www.information-management.com/newsletters/enterprise_architecture_data_model_ERP_BI-10020246-1.html?pg=2

Data Modeling for Business Value

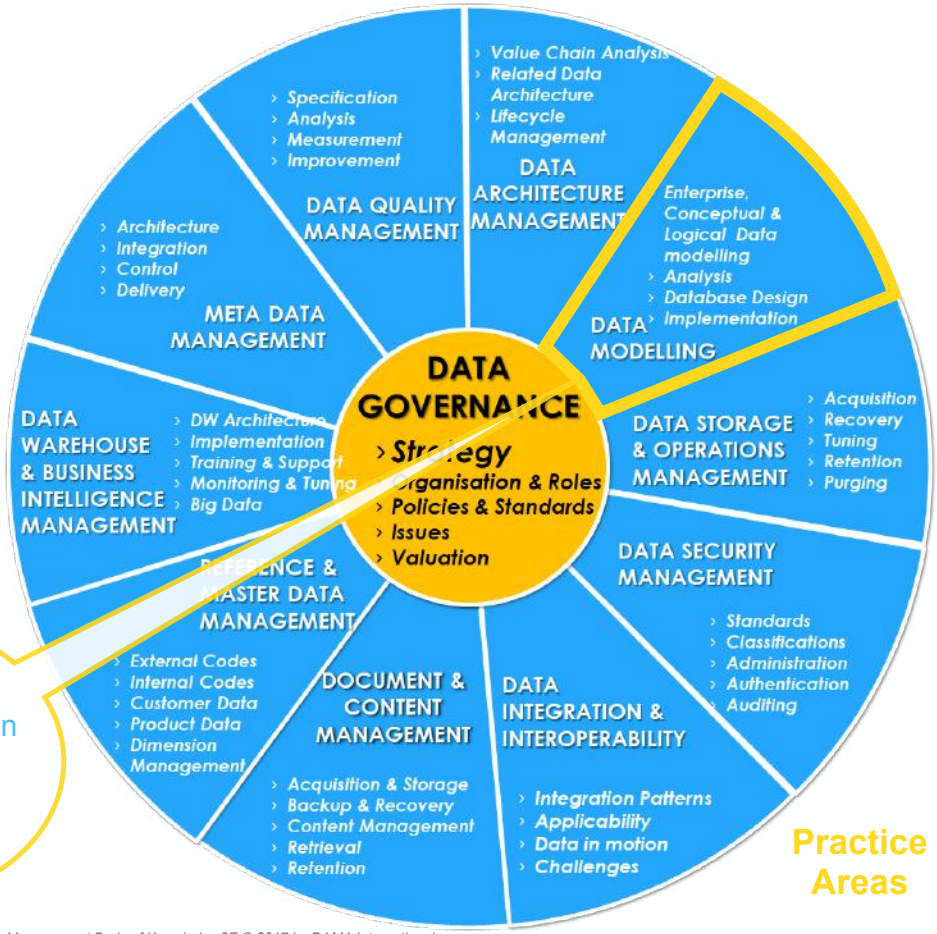
- Goal must be shared IT/business understanding
  - No disagreements/refinements means insufficient communication
- Data sharing/exchange is automated and dependent on successful engineering/architecture
  - Requires a sound foundation of data modeling basics (the essence) on which to build technologies
- Incorporate motivation (purpose statements) in all modeling
  - Modeling is a problem defining as well as a problem solving activity
- Modeling characteristics evolve during the analysis
  - Different modeling challenges for different problems
  - Use of modeling is more important than use of a specific method
  - Models must be maintained as living documents
  - 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 interaction
- Value is derived from
  - Improving organizational data
  - Improving the way people use data
  - Improving peoples use of data to support strategy

This can only be accomplished incrementally using an iterative, approach focusing on one aspect at a time and applying formal transformation methods



To Learn More

Data Management Body of Knowledge (DM BoK V2)



- Analysis
- Database Design
- Implementation
- Additional data development

Practice Areas

from *The DAMA Guide to the Data Management Body of Knowledge* 2E © 2017 by DAMA International

Works Cited

6. Works Cited / Recommended

Ambler, Scott. *Agile Database Techniques: Effective Strategies for the Agile Software Developer*. Wiley and Sons, 2003. Print.

Avison, David and Christine Cuthbertson. *A Management Approach to Database Applications*. McGraw-Hill Publishing Co., 2002. Print. Information systems ser.

Blaha, Michael. *UML Database Modeling Workbook*. Technics Publications, LLC, 2013. Print.

Brackett, Michael H. *Data Resource Design: Reality Beyond Illusion*. Technics Publications, LLC, 2012. Print.

Brackett, Michael H. *Data Resource Integration: Understanding and Resolving a Disparate Data Resource*. Technics Publications, LLC, 2012. Print.

Brackett, Michael H. *Data Resource Simplexity: How Organizations Choose Data Resource Success or Failure*. Technics Publications, LLC, 2011. Print.

Bruce, Thomas A. *Designing Quality Databases with IDEF1X Information Models*. Dorset House, 1991. Print.

Burns, Larry. *Building the Agile Database: How to Build a Successful Application Using Agile Without Sacrificing Data Management*. Technics Publications, LLC, 2011. Print.

Carlis, John and Joseph Maguire. *Mastering Data Modeling - A User-Driven Approach*. Addison-Wesley Professional, 2000. Print.

Works Cited

Codd, Edward F. "A Relational Model of Data for Large Shared Data Banks". *Communications of the ACM*, 13, No. 6 (June 1970).

DAMA International. *The DAMA Dictionary of Data Management. 2nd Edition: Over 2,000 Terms Defined for IT and Business Professionals*. 2nd ed. Technics Publications, LLC, 2011. Print.

Daoust, Norman. *UML Requirements Modeling for Business Analysts: Steps to Modeling Success*. Technics Publications, LLC, 2012. Print.

Date, C. J. *An Introduction to Database Systems*. 8th ed. Addison-Wesley, 2003. Print.

Date, C. J. and Hugh Darwen. *Databases, Types and the Relational Model*. 3d ed. Addison Wesley, 2006. Print.

Date, Chris J. *The Relational Database Dictionary: A Comprehensive Glossary of Relational Terms and Concepts, with Illustrative Examples*. O'Reilly Media, 2006. Print.

Dorsey, Paul. *Enterprise Data Modeling Using UML*. McGraw-Hill Osborne Media, 2009. Print.

Edvinsson, Håkan and Lottie Aderinne. *Enterprise Architecture Made Simple: Using the Ready, Set, Go Approach to Achieving Information Centricity*. Technics Publications, LLC, 2013. Print.

Fleming, Candace C. and Barbara Von Halle. The Handbook of Relational Database Design. Addison Wesley, 1989. Print.

Giles, John. *The Nimble Elephant: Agile Delivery of Data Models using a Pattern-based Approach*. Technics Publications, LLC, 2012. Print.

Golden, Charles. *Data Modeling 152 Success Secrets - 152 Most Asked Questions On Data Modeling - What You Need to Know*. Emereo Publishing, 2015. Print. Success Secrets.

Halpin, Terry, Ken Evans, Pat Hallock, and Bill McLean. *Database Modeling with Microsoft Visio for Enterprise Architects*. Morgan Kaufmann, 2003. Print. The Morgan Kaufmann Series in Data Management Systems.

Halpin, Terry. *Information Modeling and Relational Databases*. Morgan Kaufmann, 2001. Print. The Morgan Kaufmann Series in Data Management Systems.

Halpin, Terry. *Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design*. Morgan Kaufmann, 2001. Print. The Morgan Kaufmann Series in Data Management Systems.

Harrington, Jan L. *Relational Database Design Clearly Explained*. 2nd ed. Morgan Kaufmann, 2002. Print. The Morgan Kaufmann Series in Data Management Systems.

Hay, David C. *Data Model Patterns: A Metadata Map*. Morgan Kaufmann, 2006. Print. The Morgan Kaufmann Series in Data Management Systems.

Hay, David C. *Enterprise Model Patterns: Describing the World (UML Version)*. Technics Publications, LLC, 2011. Print.

Hay, David C. *Requirements Analysis from Business Views to Architecture*. Prentice Hall, 2002. Print.

Hay, David C. *UML and Data Modeling: A Reconciliation*. Technics Publications, LLC, 2011. Print.

Hernandez, Michael J. *Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design*. 2nd ed. Addison-Wesley Professional, 2003. Print.

Hoberman, Steve, Donna Burbank, Chris Bradley, et al. *Data Modeling for the Business: A Handbook for Aligning the Business with IT using High-Level Data Models*. Technics Publications, LLC, 2009. Print. Take It with You Guides.

Hoberman, Steve. *Data Model Scorecard*. Technics Publications, LLC, 2015. Print.

Hoberman, Steve. *Data Modeling Made Simple with ER/Studio Data Architect*. Technics Publications, LLC, 2013. Print.

Hoberman, Steve. *Data Modeling Made Simple: A Practical Guide for Business and IT Professionals*. 2nd ed. Technics Publications, LLC, 2009. Print.

Works Cited

Hoberman, Steve. *Data Modeling Master Class Training Manual*. 7th ed. Technics Publications, LLC, 2017. Print.

Hoberman, Steve. *The Data Modeler's Workbench. Tools and Techniques for Analysis and Design*. Wiley, 2001. Print.

Hoffer, Jeffrey A., Joey F. George, and Joseph S. Valacich. *Modern Systems Analysis and Design*. 7th ed. Prentice Hall, 2013. Print.

IIBA and Kevin Brennan, ed. *A Guide to the Business Analysis Body of Knowledge (BABOK Guide). International Institute of Business Analysis*, 2009. Print.

Kent, William. *Data and Reality: A Timeless Perspective on Perceiving and Managing Information in Our Imprecise World*. 3d ed. Technics Publications, LLC, 2012. Print.

Krogstie, John, Terry Halpin, and Keng Siau, eds. *Information Modeling Methods and Methodologies: Advanced Topics in Database Research*. Idea Group Publishing, 2005. Print. Advanced Topics in Database Research.

Linstedt, Dan. *Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault*. Amazon Digital Services. 2012. Data Warehouse Architecture Book 1.

Muller, Robert J. *Database Design for Smarties: Using UML for Data Modeling*. Morgan Kaufmann, 1999. Print. The Morgan Kaufmann Series in Data Management Systems.

Needham, Doug. *Data Structure Graphs: The structure of your data has meaning*. Doug Needham Amazon Digital Services, 2015. Kindle.

Newton, Judith J. and Daniel Wahl, eds. *Manual for Data Administration*. NIST Special Publications, 1993. Print.

Pascal, Fabian. *Practical Issues in Database Management: A Reference for The Thinking Practitioner*. Addison-Wesley Professional, 2000. Print.

Reingruber, Michael. C. and William W. Gregory. *The Data Modeling Handbook: A Best-Practice Approach to Building Quality Data Models*. Wiley, 1994. Print.

Riordan, Rebecca M. *Designing Effective Database Systems*. Addison-Wesley Professional, 2005. Print.

Roh, Peter and Carlos Coronel. *Database Systems: Design, Implementation, and Management*. 7th ed. Cengage Learning, 2006. Print.

Schmidt, Bob. *Data Modeling for Information Professionals*. Prentice Hall, 1998. Print.

Silverston, Len and Paul Agnew. *The Data Model Resource Book, Volume 3: Universal Patterns for Data Modeling*. Wiley, 2008. Print.

Silverston, Len. *The Data Model Resource Book, Volume 1: A Library of Universal Data Models for All Enterprises*. Rev. ed. Wiley, 2001. Print.

Silverston, Len. *The Data Model Resource Book, Volume 2: A Library of Data Models for Specific Industries*. Rev. ed. Wiley, 2001. Print.

Simsion, Graeme C. and Graham C. Witt. *Data Modeling Essentials*. 3rd ed. Morgan Kaufmann, 2004. Print.

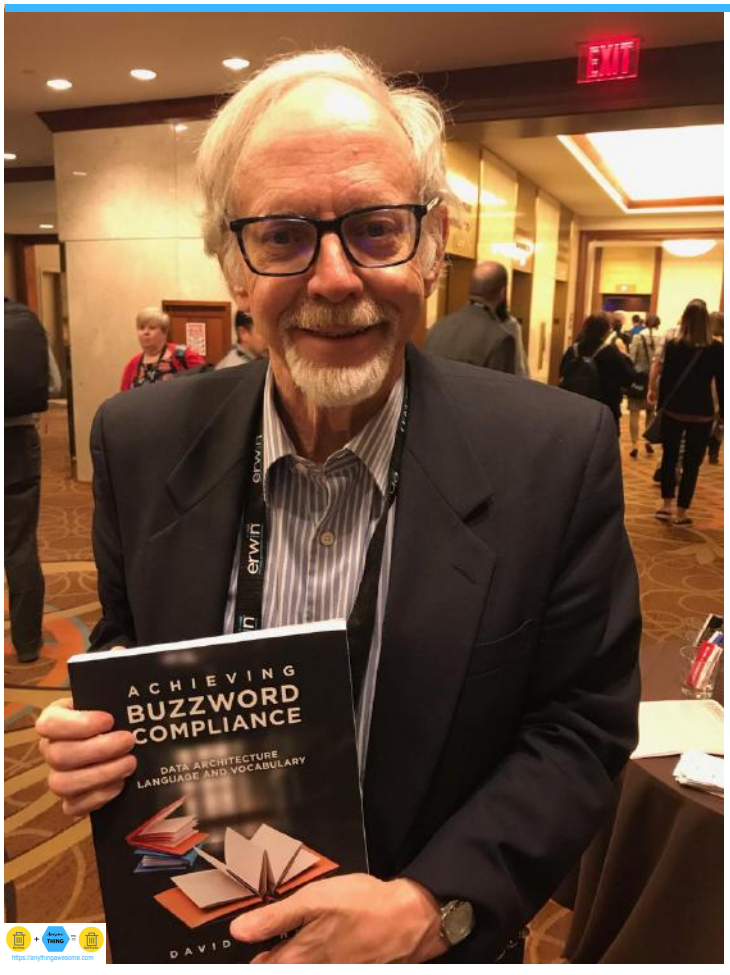
Simsion, Graeme. *Data Modeling: Theory and Practice*. Technics Publications, LLC, 2007. Print.

Teorey, Toby, et al. *Database Modeling and Design: Logical Design*, 4th ed. Morgan Kaufmann, 2010. Print. The Morgan Kaufmann Series in Data Management Systems.

Thalheim, Bernhard. *Entity-Relationship Modeling: Foundations of Database Technology*. Springer, 2000. Print.

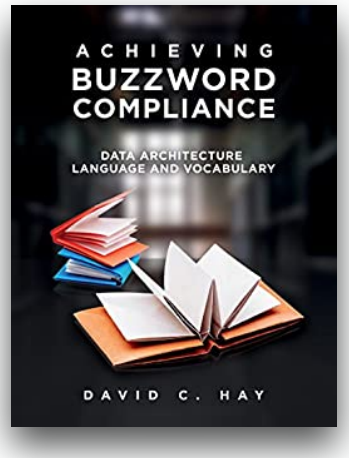
Watson, Richard T. *Data Management: Databases and Organizations*. 5th ed. Wiley, 2005. Print.

from *The DAMA Guide to the Data Management Body of Knowledge* 2E © 2017 by DAMA International



Achieving Buzzword Compliance

Data Architecture Language and Vocabulary

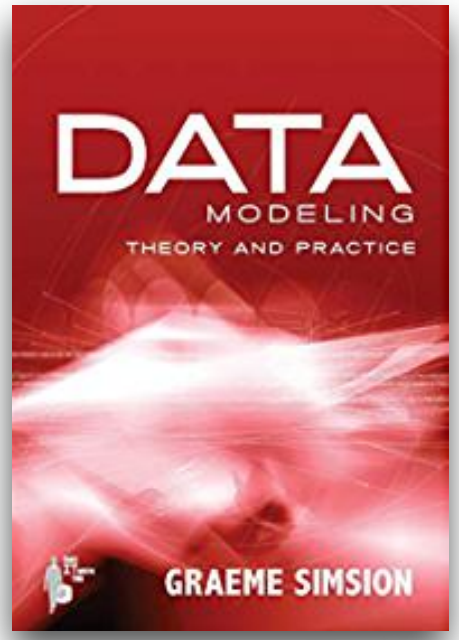


amazon.com link:

https://www.amazon.com/Achieving-Buzzword-Compliance-Architecture-Vocabulary-ebook/dp/B07FG1WRSD/ref=sr_1_1?crid=2QL3ZWKU2L3VC&keywords=Achieving+Buzzword+Compliance%3A+Data+Architecture+Language+and+Vocabulary&qid=1657032460&sprefix=achieving+buzzword+compliance+data+architecture+language+and+vocabulary%2Caps%2C324&sr=8-1

Data Modeling: Theory and Practice

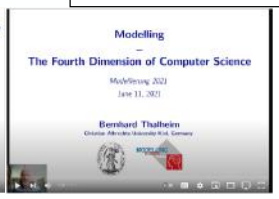
Graeme Simsion



Research Efforts

- Professor Bernhard Thalheim and associated research efforts have contributed much to these topics including:

- Conceptual modelling
  - https://www.youtube.com/watch?v=Y9_7KSsSUpg
  - https://www.youtube.com/watch?v=mKcwbR6uJwU



- Claim: logical models also conceptual models
  - https://www.youtube.com/watch?v=L8yGjEbwTsQ
  - https://link.springer.com/article/10.1007/s10270-020-00836-z

https://www.youtube.com/watch?v=9qWjpVtr_Hg&t=2s

Advanced Data Modeling class 2016

GETITLE

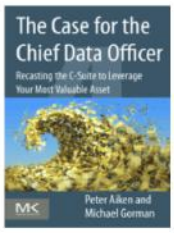
Conceptual vs. Logical vs. Physical

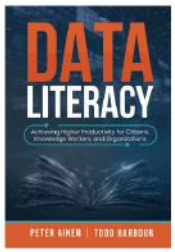
Stages of Data Modeling

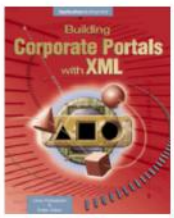
©Gordon C. Everest
Professor Emeritus of MIS and Database
Carlson School of Management
University of Minnesota
geverest@umn.edu http://geverest.umn.edu

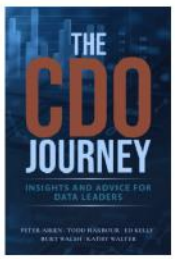
Event Pricing on Peter's Books

- 20% off directly from the publisher on select titles
- My 'Book Store' @ https://anythingawesome.com/books-overview.html
- Enter the code "anythingawesome" at the Technics bookstore checkout where it says to "Apply Coupon"









[Clicking any webinar title will link directly to the registration page]

Upcoming Events

The Importance of Metadata: 3 Leveraging Strategies

8 Aug 2023

Time: 19:00 UTC (2:00 PM NYC) | Presented by: Peter Aiken, PhD

Getting Data Quality Right

12 September 2023

ENTERPRISE DATA WORLD

Strategy Is Where Data Architecture and Data Governance Collide

10 October 2023

Brought to you by:

DATAVERSITY

Critical Design Review?

Mentoring?

Collaboration?

Executive Data Literacy Training?

Independent Verification & Validation

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - https://anythingawesome.com/OfficeHours.html