



EDUCATION

LEARN AND LEARN

TRAINING

LEADERSHIP

SUCCESS

# Data Modelling 101

CHRISTOPHER BRADLEY

DATA MANAGEMENT ADVISORS



@InfoRacer

T: +44 7973 184475

E: Chris.Bradley@DMAdivisors.co.uk

**Blog:** Information Management, Life & Petrol  
<http://infomanagementlifeandpetrol.blogspot.com>

LI: [uk.linkedin.com/in/christophermichaelbradley/](http://uk.linkedin.com/in/christophermichaelbradley/)

# Christopher Bradley

INFORMATION MANAGEMENT STRATEGIST





Enterprise Data & BI  
Conference Europe 2014

**3-5 November 2014**

Produced by:





# Christopher Bradley

Chris has 35 years of Information Management experience & is a leading Independent Information Management strategy advisor.

In the Information Management field, Chris works with prominent organizations including Vodafone, BT, HSBC, Celgene, GSK, Pfizer, Icon, Quintiles, Total, Barclays, ANZ, GSK, Shell, BP, Statoil, Riyadh Bank & Aramco. He addresses challenges faced by large organisations in the areas of Data Governance, Master Data Management, Information Management Strategy, Data Quality, Metadata Management and Business Intelligence.

He is a Director of DAMA- I, holds the CDMP Master certification, is an examiner for CDMP, a Fellow of the Chartered Institute of Management Consulting (now IC)

a member of the MPO, and SME Director of the DM Board.

A recognised thought-leader in Information Management Chris is the author of numerous papers, books, including sections of DMBok 2.0, a columnist, a frequent contributor to industry publications and member of several IM standards authorities.

He leads an experts channel on the influential BeyeNETWORK, is a sought after speaker at major international conferences, and is the co-author of "Data Modelling For The Business – A Handbook for aligning the business with IT using high-level data models". He also blogs frequently on Information Management (and motorsport).



**MELTON**

CONCRETE REINFORCEMENT

REINFORCING AND PRECAST  
CONCRETE FLOORS AND STAIRS

for the construction industry

01209 863288



# Recent Presentations

**Enterprise Data World:** (DataVersity), May 2014, Austin, Texas, "MDM Architectures & How to identify the right Subject Area & tooling for your MDM strategy"

**E&P Information Management Dubai:** (DMBoard), 17-19 March 2014, Dubai, UAE  
"Master Data Management Fundamentals, Architectures & Identify the starting Data Subject Areas"

**DAMA Australia:** (DAMA-A), 18-21 November 2013, Melbourne, Australia "DAMA DMBok 2.0", "Information Management Fundamentals" 1 day workshop"

**Data Management & Information Quality Europe:** (IRM Conferences), 4-6 November 2013, London, UK  
"Data Modelling Fundamentals" ½ day workshop:

"Myths, Fairy Tales & The Single View" Seminar

"Imaginative Innovation - A Look to the Future" DAMA Panel Discussion

**IPL / Embarcadero series:** June 2013, London, UK, "Implementing Effective Data Governance"

**Riyadh Information Exchange:** May 2013, Riyadh, Saudi Arabia,  
"Big Data – What's the big fuss?"

**Enterprise Data World:** (Wilshire Conferences), May 2013, San Diego, USA, "Data and Process Blueprinting – A practical approach for rapidly optimising Information Assets"

**Data Governance & MDM Europe:** (IRM Conferences), April 2013, London, "Selecting the Optimum Business approach for MDM success.... Case study with Statoil"

**E&P Information Management:** (SMI Conference), February 2013, London,  
"Case Study, Using Data Virtualisation for Real Time BI & Analytics"

**E&P Data Governance:** (DMBoard / DG Events), January 2013, Marrakech, Morocco,  
"Establishing a successful Data Governance program"

**Big Data 2:** (Whitehall), December 2012, London, "The Pillars of successful knowledge management"

**Financial Information Management Association (FIMA):** (WBR), November 2012, London; "Data Strategy as a Business Enabler"

**Data Modeling Zone:** (Technics), November 2012, Baltimore USA  
"Data Modelling for the business"

**Data Management & Information Quality Europe:** (IRM), November 2012, London; "All you need to know to prepare for DAMA CDMP professional certification"

**ECIM Exploration & Production:** September 2012, Haugesund, Norway:  
"Enhancing communication through the use of industry standard models; case study in E&P using WITSML"

**Preparing the Business for MDM success:** Threadneedles Executive breakfast briefing series, July 2012, London

**Big Data – What's the big fuss?:** (Whitehall), Big Data & Analytics, June 2012, London,

**Enterprise Data World International:** (DAMA / Wilshire), May 2012, Atlanta GA,  
"A Model Driven Data Governance Framework For MDM - Statoil Case Study"  
"When Two Worlds Collide – Data and Process Architecture Synergies" (*rated best workshop in conference*); "Petrochemical Information Management utilising PPDM in an Enterprise Information Architecture"

**Data Governance & MDM Europe:** (DAMA / IRM), April 2012, London,  
"A Model Driven Data Governance Framework For MDM - Statoil Case Study"

**AAPG Exploration & Production Data Management:** April 2012, Dead Sea Jordan; "A Process For Introducing Data Governance into Large Enterprises"

**PWC & Iron Mountain Corporate Information Management:** March 2012, Madrid;  
"Information Management & Regulatory Compliance"

**DAMA Scandinavia:** March 2012, Stockholm,  
"Reducing Complexity in Information Management" (*rated best presentation in conference*)

**Ovum IT Governance & Planning:** March 2012, London;  
"Data Governance – An Essential Part of IT Governance"

**American Express Global Technology Conference:** November 2011, UK,  
"All An Enterprise Architect Needs To Know About Information Management"

**FIMA Europe (Financial Information Management):**, November 2011, London; "Confronting The Complexities Of Financial Regulation With A Customer Centric Approach; Applying IPL's Master Data Management And Data Governance Process In Clydesdale Bank "

**Data Management & Information Quality Europe:** (DAMA / IRM), November 2011, London,  
"Assessing & Improving Information Management Effectiveness – Cambridge University Press Case Study"; "Too Good To Be True? – The Truth About Open Source BI"

**ECIM Exploration & Production:** September 12th 14th 2011, Haugesund, Norway: "The Role Of Data Virtualisation In Your EIM Strategy"

**Enterprise Data World International:** (DAMA / Wilshire), April 2011, Chicago IL; "How Do You Want Yours Served? – The Role Of Data Virtualisation And Open Source BI"

**Data Governance & MDM Europe:** (DAMA / IRM), March 2011, London,  
"Clinical Information Data Governance"

**Data Management & Information Management Europe:** (DAMA / IRM), November 2010, London,  
"How Do You Get A Business Person To Read A Data Model?"

**DAMA Scandinavia:** October 26<sup>th</sup>-27<sup>th</sup> 2010, Stockholm,  
"Incorporating ERP Systems Into Your Overall Models & Information Architecture" (*rated best presentation in conference*)

**BPM Europe:** (IRM), September 27<sup>th</sup> – 29<sup>th</sup> 2010, London,  
"Learning to Love BPMN 2.0"

**IPL / Composite Information Management in Pharmaceuticals:** September 15<sup>th</sup> 2010, London,  
"Clinical Information Management – Are We The Cobblers Children?"

**ECIM Exploration & Production:** September 13<sup>th</sup> 15<sup>th</sup> 2010, Haugesund, Norway: "Information Challenges and Solutions" (*rated best presentation in conference*)

**Enterprise Architecture Europe:** (IRM), June 16<sup>th</sup> – 18<sup>th</sup> 2010, London: ½ day workshop; "The Evolution of Enterprise Data Modelling"

# Recent Publications

- Book:** "Data Modelling For The Business – A Handbook for aligning the business with IT using high-level data models"; Technics Publishing; ISBN 978-0-9771400-7-7; <http://www.amazon.com/Data-Modeling-Business-Handbook-High-Level>
- White Paper:** ["Information is at the heart of ALL Architecture disciplines"](#); March 2014
- Article:** [The Bookbinder, the Librarian & a Data Governance story](#) ; July 2013
- Article:** [Data Governance is about Hearts and Minds, not Technology](#) January 2013
- White Paper:** ["The fundamentals of Information Management"](#), January 2013
- White Paper:** ["Knowledge Management – From justification to delivery"](#), December 2012
- Article:** "Chief INFORMATION Officer? Not really" [Article](#), November 2012
- White Paper:** "Running a successful Knowledge Management Practice" November 2012
- White Paper:** ["Big Data Projects are not one man shows"](#) June 2012
- Article:** "IPL & Statoil's innovative approach to Master Data Management in Statoil", [Oil IT Journal](#), May 2012
- White Paper:** ["Data Modelling is NOT just for DBMS's"](#) April 2012
- Article:** "Data Governance in the Financial Services Sector" [FSTech Magazine](#), April 2012
- Article:** ["Data Governance, an essential component of IT Governance"](#) March 2012
- Article:** "Leveraging a Model Driven approach to Master Data Management in Statoil", [Oil IT Journal](#), February 2012
- Article:** "How Data Virtualization Helps Data Integration Strategies" [BeyeNETWORK \(December 2011\)](#)
- Article:** "Approaches & Selection Criteria For organizations approaching data integration programmes" [TechTarget \(November 2011\)](#)
- Article:** Big Data – Same Problems? [BeyeNETWORK](#) and [TechTarget](#). (July 2011)
- Article** "10 easy steps to evaluate Data Modelling tools" [Information Management](#), (March 2010)
- Article** "How Do You Want Your Data Served?" [Conspectus Magazine](#) (February 2010)
- Article** "How do you want yours served (data that is)" ([BeyeNETWORK January 2010](#))
- Article** "Seven deadly sins of data modelling" ([BeyeNETWORK October 2009](#))
- Article** "Data Modelling is NOT just for DBMS's" [Part 1 BeyeNETWORK July 2009](#) and [Part 2 BeyeNETWORK August 2009](#)
- Web Channel:** BeyeNETWORK "Chris Bradley Expert Channel" Information Asset Management <http://www.b-eye-network.co.uk/channels/1554/>
- Article:** ["Preventing a Data Disaster"](#) February 2009, Database Marketing Magazine



# Information Management Training & Mentoring

We offer a number of training options & **Custom-built**, on-site training & awareness seminars can also be delivered.

The following training courses are available:



- **Information Management Fundamentals – 5 day** introductory course covering all of the components of Information Management as defined in the DAMA Body of Knowledge (DMBoK) & forthcoming changes in DMBoK 2.0 presented by one of the DMBoK 2 authors
- **Data Modelling fundamentals – 3 day** intermediate course introducing students to data modelling, its purpose, the different types of models and how to construct and read a data model.
- **Advanced Data Modeling – 3 day** advanced course for students with data modelling experience to understand the advanced concepts and human centric aspects of data modelling to enable them to build quality models that meet business needs.
- **IM Fundamentals & Practitioner Courses**– A series of **1 day** (*foundation*) and **2 day** (*practitioner*) classes to give practitioners a solid background in a specific Information Management topics. The 2 day practitioner workshops explore more detail on the implementation aspects of the particular Information Management discipline
  - Data Modelling Foundation (*1 day only*)
  - Data Governance Foundation & Practitioner
  - Master & Reference Data Foundation & Practitioner
  - Data Quality Management Foundation & Practitioner
  - Data Warehouse & Business Intelligence Foundation & Practitioner
  - Data Integration Foundation & Practitioner
- **Executive Workshops – ½ and 1 day** executive workshop(s) designed to give non-technical managers a basic understanding of a various Information Management topics and their importance to the organisation.
- **CDMP Certification– 3 day** workshop “exam cram” designed to help attendees pass the DAMA CDMP certification. Sitting the live examinations is included as part of the workshop.
- **Integrated Business Process, Data & Requirements Definition– 5 day** intensive class to show students an integrated requirements discovery and definition approach covering business process, different types of requirements modelling, and the critical role of the conceptual data model.



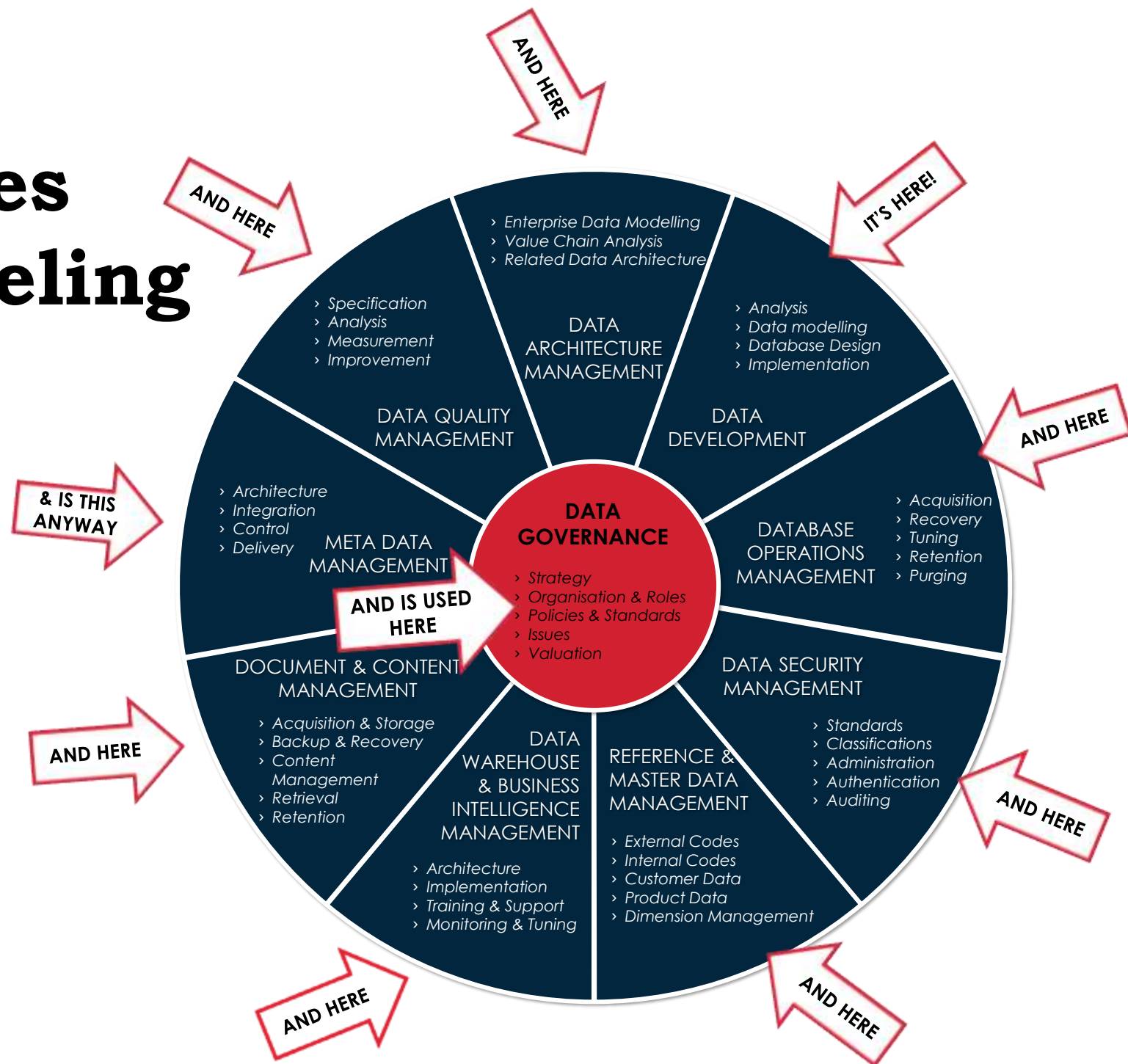


# *Data Modelling 101*

# Data Modelling 101

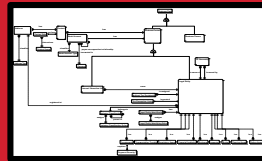
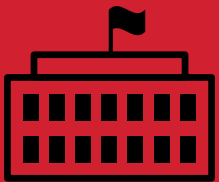
- › **CONTEXT WITHIN THE DMBOK**
- › **DATA & METADATA**
- › **DATA MODELLING: WHAT & WHY?**
- › **TYPES & LEVELS OF DATA MODELS**
- › **DATA MODEL COMPONENTS**
- › **NORMALISATION**
- › **DIMENSIONAL DATA MODELLING**
- › **IT'S NOT JUST FOR DBMS'S**
- › **SUMMARY**

# Where does Data Modeling fit?

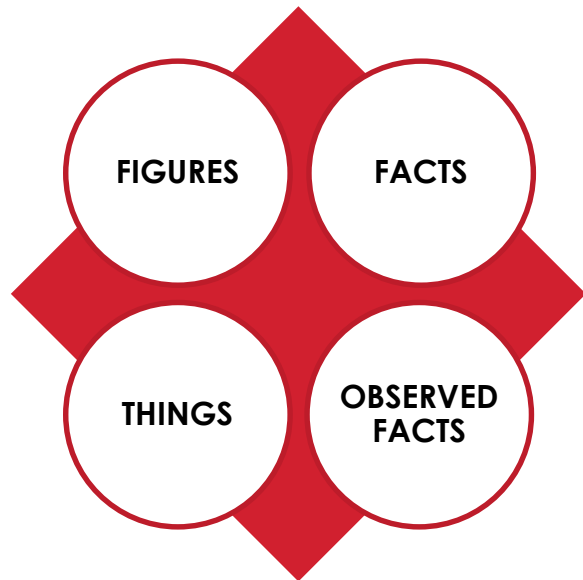




# Is there more to life than this?



# What is Data?




*Data in context = Information*

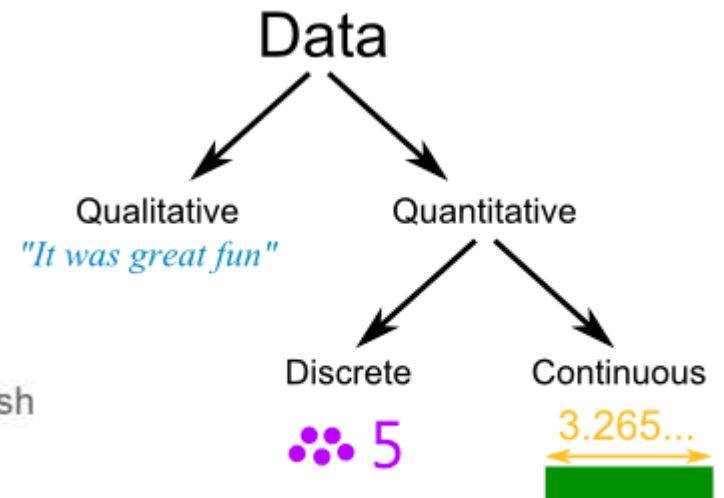
**Current Balance** = \$400

**Home Town** = Bath

**Order Placed Date** = 4<sup>th</sup> November 2013

**Customer id** = 987654321

**Person image** = 



## Definition of data in English

data

Pronunciation: /ˈdeɪtə/

Translate data | into French | into German | into Italian | into Spanish

noun

[mass noun]

facts and statistics collected together for reference or analysis:

*there is very little data available*

- the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
- *Philosophy* things known or assumed as facts, making the basis of reasoning or calculation.

Origin:

mid 17th century (as a term in philosophy): from Latin, plural of DATUM

In Latin, **data** is the plural of **datum** and, historically and in specialized scientific fields, it is also treated as a plural in English, taking a plural verb, as in *the data were collected and classified*. In modern non-scientific use, however, it is generally not treated as a plural. Instead, it is treated as a mass noun, similar to a word like **information**, which takes a singular verb. Sentences such as *data was collected over a number of years* are now widely accepted in standard English.

*Data is objective*

# Data

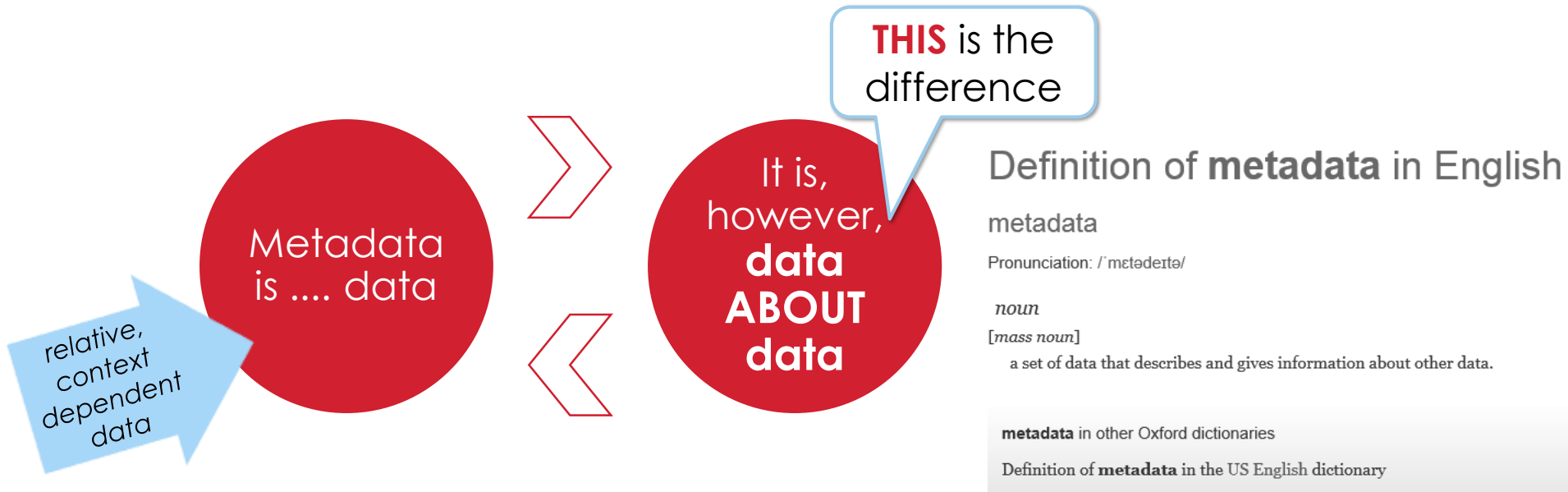
Person id	Forename	Family Name	Salutation	NI number	Player Rating	Office address	Credit limit	Emergency Contact
12356	Mitchell	Stark	Mr	112113	2	123 St James Place, London, WC1	\$0.25m	F.J Banks
124	Gary	Sobers	Sir	112141	2	Shellmex house, London, EC1	\$0.25m	E.C. Dollar
09211	Alan	Knott	Mr	201221	4	IBM House, White Plains, NY	\$0.35m	F.E. Goodwin
43219	Rachel	Hahoe-Flint	Mrs	202119	5	Microsoft, MS Business Park, Seattle, WA	\$0.55m	R.B. Gibb
12	Allan	Border	Mr	311456	5	Dell park, Palo Alto, CA	\$0.5m	S.T.Law
230	Ian	Botham	Sir	429876	0	Seattle Aero Park, Seattle, WA	-	-



# What is Metadata?



# What is Metadata?



**Person id:** This is the unique identification number for the customer as used in our organisation.

**Forename:** The preferred name used by the person. **Note** this is **not** the same as the birth certificate forename.

**NI Number:** The National Insurance number for the person.

*Metadata is **context dependent***

**KEY** point

# Metadata

*Metadata has  
“properties”*

*These describe the  
characteristics & rules  
of the metadata*

**Description** = This  
is the unique  
identification  
number for the  
person as used  
in our  
organisation.

**Size** = 1-5  
numeric

**PERSON ID**

**Mandatory** =  
Yes

**Unique key** =  
Yes



# Where do you encounter metadata every day?



# MetaData

## DATA



## METADATA

General Permissions Meta Info Preview

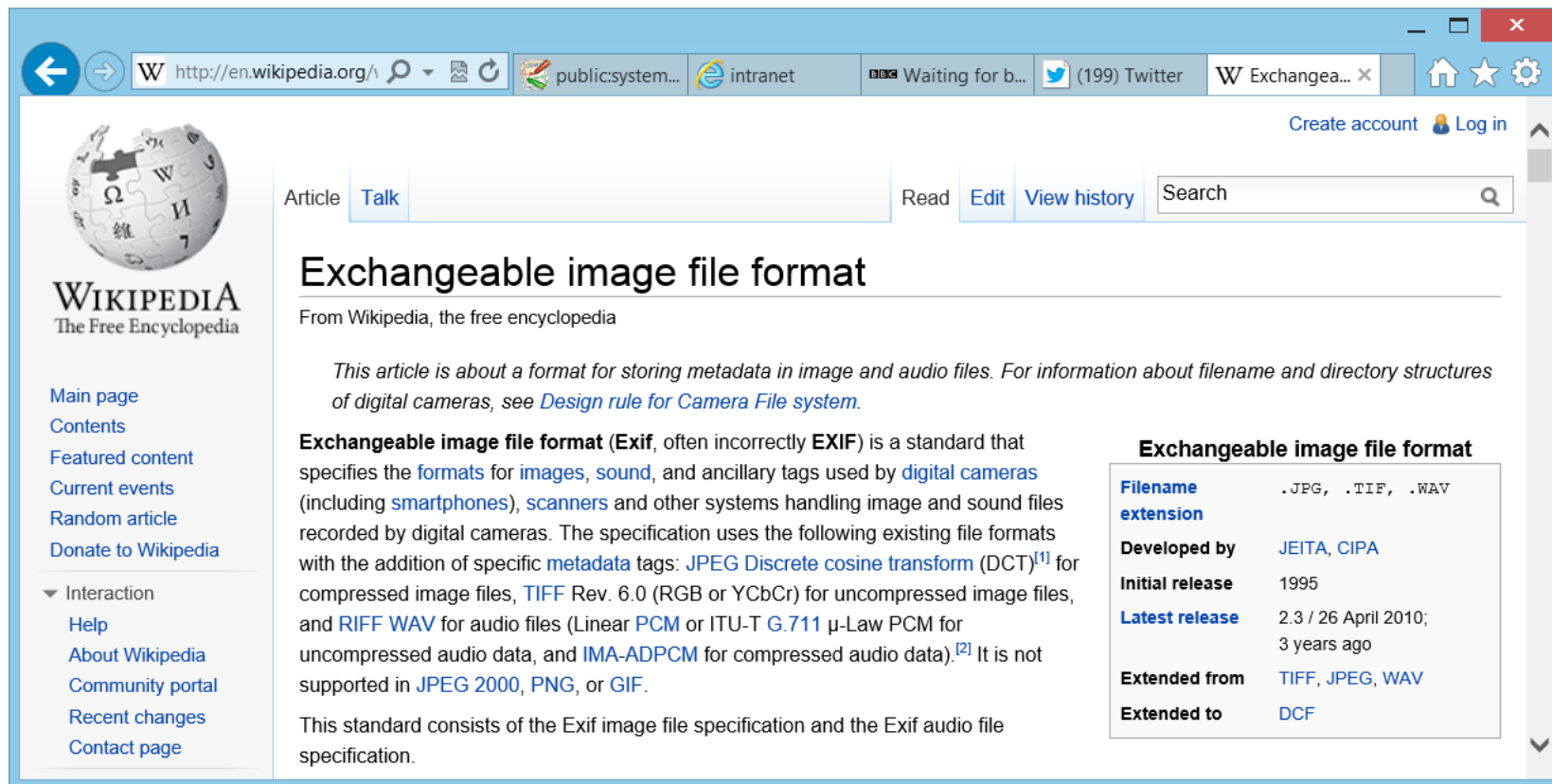
JPEG Exif

Comment:

Creation Date: 05-01-14  
Creation Time: 12:38:36 am  
Dimensions: 2560 x 1920 pixels  
Exposure Time: 0.100 (1/10)  
JPEG Quality: Unknown  
Aperture: f/3.3  
Color Mode: Color  
Date/Time: 05-01-14 12:38:36 am  
Flash Used: Off  
Focal Length: 6.3 mm  
ISO Equiv.: 100  
JPEG Process: Baseline  
Camera Manufacturer: PENTAX Corporation  
Metering Mode: Pattern  
Camera Model: PENTAX Optio WP  
Orientation: 1

OK Cancel

# MetaData



← → W http://en.wikipedia.org/ publicsystem... intranet Waiting for b... (199) Twitter Exchangea... x

Create account Log in

Article **Talk** Read Edit View history Search

## Exchangeable image file format

From Wikipedia, the free encyclopedia

*This article is about a format for storing metadata in image and audio files. For information about filename and directory structures of digital cameras, see [Design rule for Camera File system](#).*

**Exchangeable image file format (Exif**, often incorrectly **EXIF**) is a standard that specifies the [formats](#) for [images](#), [sound](#), and ancillary tags used by [digital cameras](#) (including [smartphones](#)), [scanners](#) and other systems handling image and sound files recorded by digital cameras. The specification uses the following existing file formats with the addition of specific [metadata](#) tags: [JPEG Discrete cosine transform](#) (DCT)<sup>[1]</sup> for compressed image files, [TIFF](#) Rev. 6.0 (RGB or YCbCr) for uncompressed image files, and [RIFF WAV](#) for audio files (Linear [PCM](#) or ITU-T [G.711](#)  $\mu$ -Law PCM for uncompressed audio data, and [IMA-ADPCM](#) for compressed audio data).<sup>[2]</sup> It is not supported in [JPEG 2000](#), [PNG](#), or [GIF](#).

This standard consists of the Exif image file specification and the Exif audio file specification.

### Exchangeable image file format

<b>Filename extension</b>	.JPG, .TIF, .WAV
<b>Developed by</b>	JEITA, CIPA
<b>Initial release</b>	1995
<b>Latest release</b>	2.3 / 26 April 2010; 3 years ago
<b>Extended from</b>	TIFF, JPEG, WAV
<b>Extended to</b>	DCF



# Where else do you use metadata every day?

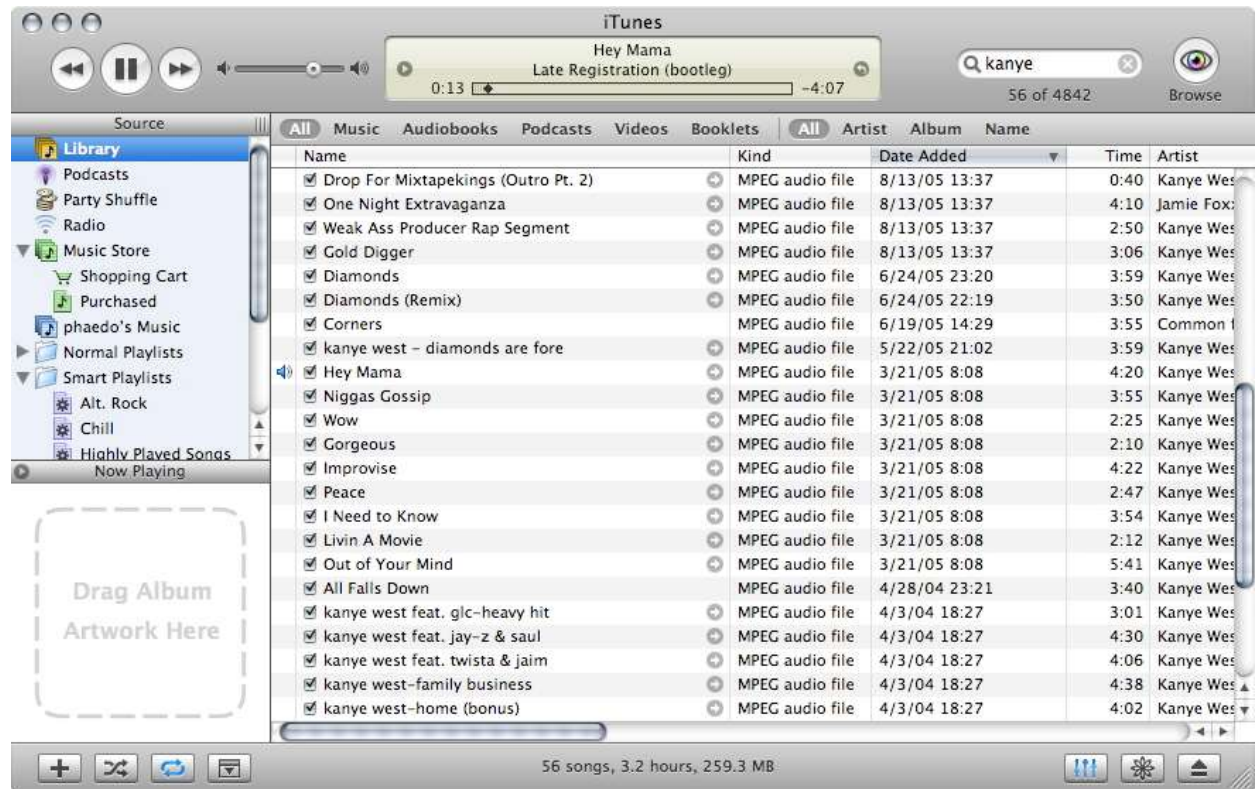


# MetaData

## DATA



## METADATA





# Where else do you use metadata every day?



Sky.com home Find & Watch TV Sign in Sign up

sky HIGHLIGHTS TV GUIDE Search Sky Guide

Now TODAY WED THU FRI SAT SUN MON ENTERTAINMENT LONDON HD Accessible TV Guide

Growing Up Hollywood Kitchen Boss LA Ink Ultimate Spider-Man To Boldly Go

2pm 1pm 2pm 3pm 4pm

one 101	Bargain Hunt Scotland: The Bargain H...	BBC News... The latest nati...	BB... The I...	Doctors Host Family: C...	Escape to the Coun... Hampshire: Nicki Chapm...	Deadly 60 CBBC, South A...	Lockie L... CBBC, Cure...	Prank P... CBBC, 14/2...	Junior Mas.. CBBC, 7/15. Fo
two 102	tics l and Jo Coburn are joi...	Restoratio... 18/20. Series...	To Buy or Not to Buy Property series. Jonny B...	The Weakest Link Anne Robinson presents...	Real Rescues 13/20. Nick Knowles and...	Flog It! Coventry: Antiques serie...	Breakaway Six players ans		
itv 1 103	Loose Women Melanie C and Emma Bunton join...	ITV News... Alastair Stewa...	Crime Stories Drama series filmed in the style of...	Dickinson's Real Deal David Dickinson and dealers David...	The Dales Adrian Edmon...	The Dales Last in series...			
4 104	crim... endin...	Come Dine with Me Debbie McGee, Lembit Opik, Iain L...	Phil Spencer: Secret Agent Phil comes to the rescue of grandmot...	Countdown Nick Hewer and Rachel Riley...	1001 Thing... Sandi Toksvig...	Deal or No Deal Noel Edmonds asks the all-importa			
5 105	est House In The Street ew series about people turning t...	Home and... When Brax fin...	Neighbours After Paul adm...	CSI: Miami Grand Prix: US drama series set in...	Bonneville Heartwarming tale starring Jessica Lange and Joan Allen. Three...				

Los Angeles NCIS: Los Angeles Glas Starski SS-1 Starski SS-1

**Exercise 1:**

# Data or Metadata

## DATA OR METADATA?

Chris Bradley

Company Name

750 metres

1 Royal Crescent, Bath



Shell

Location

Dubai

Date Of Birth

Order Date



Nov 3<sup>rd</sup> 2014

Singer Name

Order Status

Chris.bradley@DMAdvisors.co.uk

+44 7808 038 173



## Exercise 1:

# Data or Metadata

### DATA OR METADATA?

**Data** Chris Bradley

**Metadata** Date Of Birth

**Metadata** Company Name

**Metadata** Order Date

**Data** 750 metres

**Data** 

**Data** 1 Royal Crescent, Bath

**Data** Nov 3<sup>rd</sup> 2014

**Data**



**Metadata** Singer Name

**Data** Shell

**Metadata** Order Status

**Metadata** Location

**Data** Chris.bradley@DMAdvisors.co.uk

**Data** Dubai

**Data** +44 7808 038 173

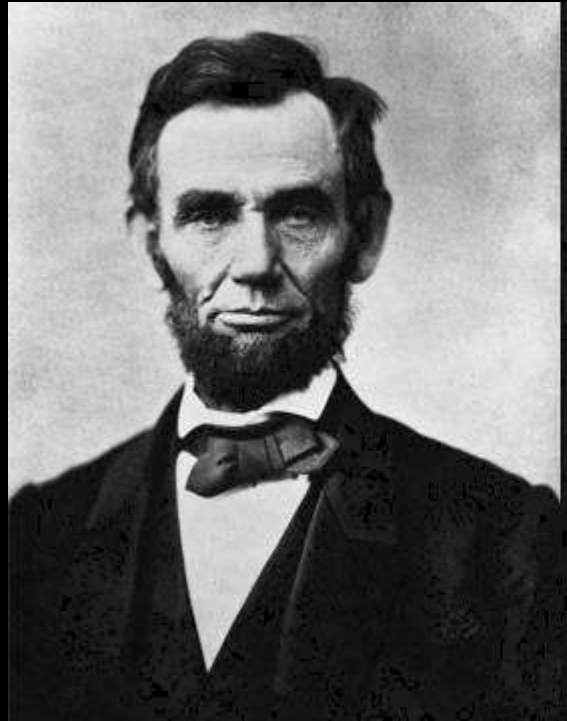
# What Is Data Modelling?

The purpose of a Data model is to design a relational database system

An ER Model is used to specify design and document Database design

A Data model is a pictorial representation of the structure of a relational database system

it is a description of the objects represented by a computer system together with their properties and relationships



**“Don’t believe everything you read on the Internet just because there’s a picture with a quote next to it.”**

—Abraham Lincoln

ER Modelling is a Database design method



# What Is A Data Model?

A **model** is a representation of something in our environment making use of standard symbols to enable improved understanding of the concept

A **data model** describes the specification, definition and rules for data in a business area

A **data model** is a diagram (with additional supporting metadata) that uses text and symbols to represent data to give the reader a better understanding of the data

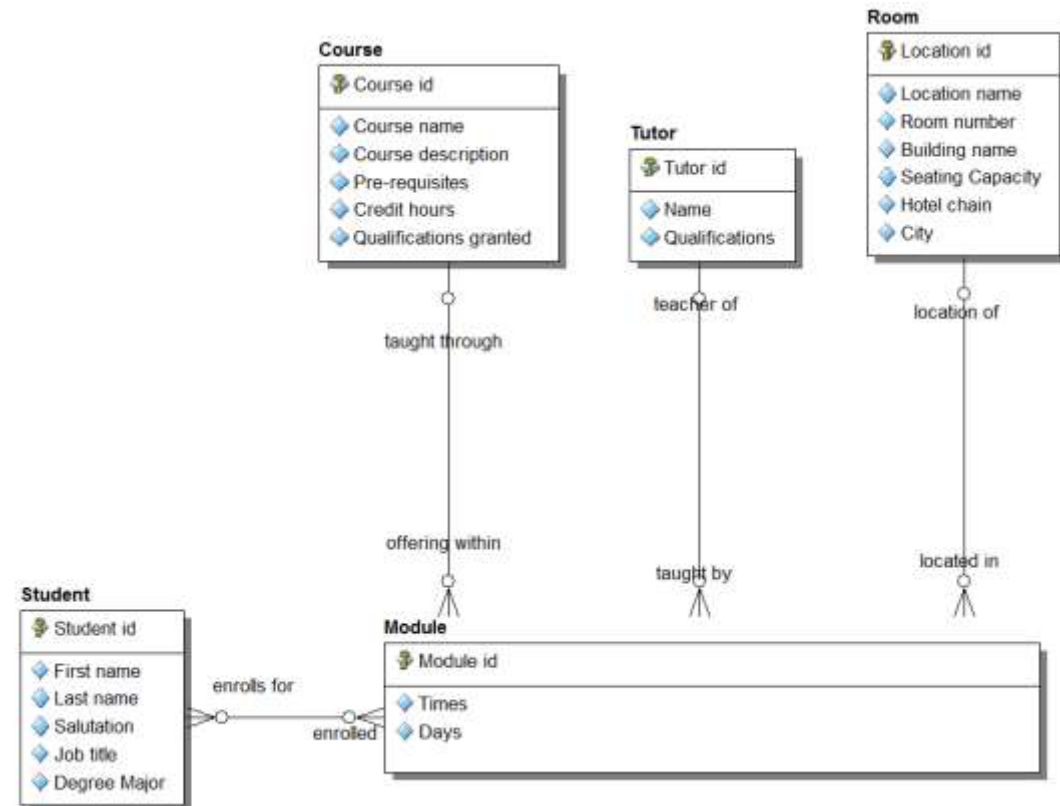
A **data model** describes the inherent logical structure of the data within a given domain and, by implication, the underlying structure of that domain itself





# What Is A Conceptual Data Model?

- A description of a Business (or an area of the Business) in terms of the **things** it needs to know about.
- The Data **things** are “entities” and the “**facts about things**” are attributes & relationships.
- It’s a representation of the “real world”, not a technical implementation of it
- *Should* be able to be understood by Business users



## Definition:

A *Student* is any person who has been admitted to a course, has paid, and has enrolled in one or more modules within a course. Tutors and other staff members may also be Students

## Business Assertions

- A Student enrolls for one or more modules
- A Course can be taught through one or more Modules
- A Room can be the location of one or more modules
- A Tutor can be the teacher of one or more modules

## The Other Way?

- A Module is enrolled in by many students
- A Module is an offering within one course
- A Module is located in one room
- A Module is taught by one tutor

} Really?

# A Data Model Represents

## WHO

*Person, Employee, Vendor, Customer,  
Department, Organisation, ...*

## WHAT

*Product, Service, Raw Material, Training  
Course, Flight, Room, ...*

## WHEN

*Time, Day, Date, Calendar, Reporting Period,  
Fiscal Period, ...*

## WHERE

*Geographic location, Delivery address,  
Storage Depot, Airport, ...*

## WHY

*Order, Complaint, Inquiry, Transaction, ...*

## HOW

*Invoice, Policy, Contract, Agreement,  
Document, Account, ...*

Classes of  
**entities**  
**(kinds of things)**  
about which a  
company  
wishes to know  
or hold  
information

# What is an Entity?

*The “Who, What, Where, When, Why” of the Organization*

**Entity:** A classification of the types of objects found in the real world -- persons, places, things, concepts and events – of interest to the enterprise. <sup>1</sup>

<sup>1</sup> DAMA Dictionary of Data Management



**WHO?**



**WHAT?**



**WHERE?**



**WHEN?**



**WHY?**



**HOW?**

# Identifying Entities

*Is it an  
Entity?*



What is **ONE** of those things?

Does this imply an instance of a **SINGLE** thing, not a group or collection

How do I identify **ONE** of those things?

What are the facts I want to hold against **ONE** of those things?

Do I even **WANT** to hold facts about these things?

**PROCESSES** will act upon it, so does the “thing” make sense in a well formed process phrase i.e. a verb – noun pair?



# Sample Entities



# Exercise 2: Entities

Which of these might / might not be valid entities?

Student

Building

Maths

Department

Course  
Catalogue

Attendance  
Sheet

Enrolment  
Form

Professor  
Plumb

Prerequisite  
list

Module

Organisation  
Chart

Student  
Directory

Module  
Description

Qualification

Certification  
Body

Graduation

# A Data Model Represents

the  
**attributes**  
of that  
information  
(**facts about  
things**)

PERSON ID

FIRST NAME

DATE OF BIRTH

PRODUCT NUMBER

QUANTITY ORDERED

FLIGHT NUMBER

SEAT CLASS

...

# Attributes

*An **Attribute** is a piece of information about or a characteristic of an Entity.*



Entity → Employee

- Employee Identifier
- Employee Last Name
- Employee First Name
- Employee Hire Date
- Employee Signed Employment Contract
- Employee Drivers License Photo

Attributes

Entity → Employee

Employee Identifier
Employee Last Name
Employee First Name
Employee Hire Date
Employee Signed Employment Contract
Employee Drivers License Photo

Attributes

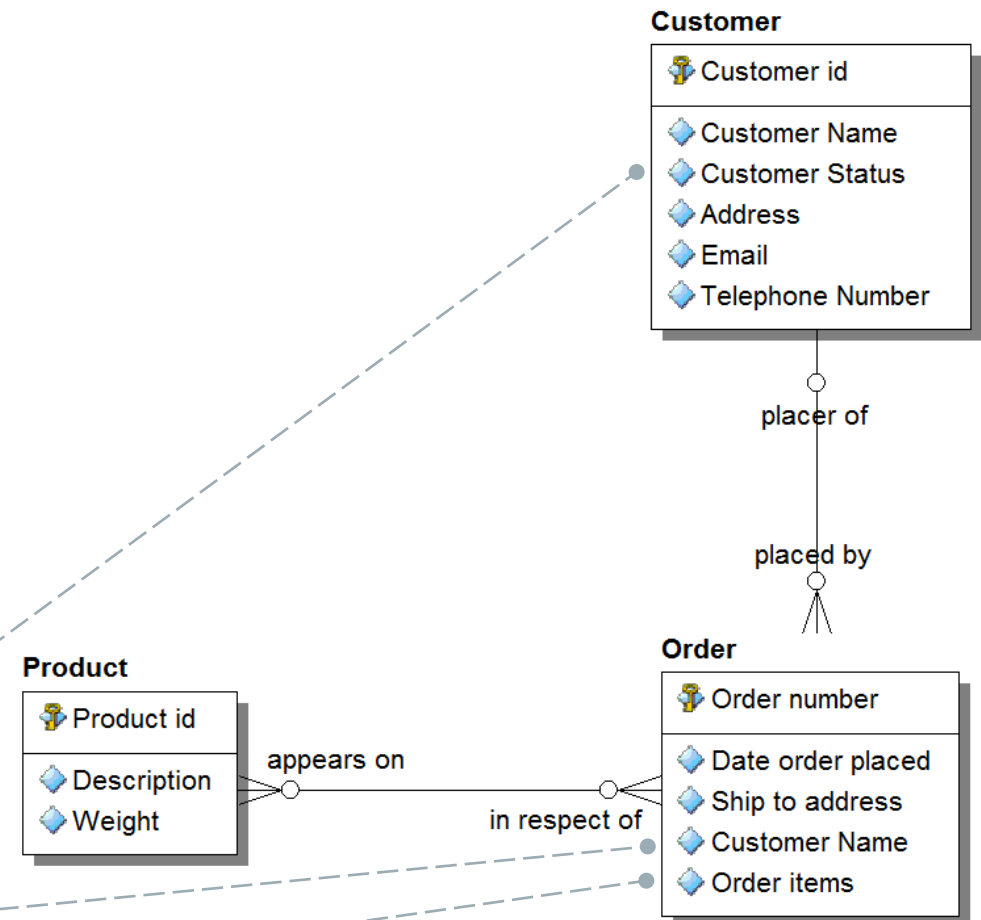
# Attributes

- Facts about “entities” are recorded as attributes & relationships.
- We don’t record *every* fact, only the ones that are needed

## Attribute Properties

- “user-entered” vs. “constrained set”: The attribute can only come from a finite set, such as code list / drop down set
- fundamental vs. redundant: the same value is recorded multiple times in different entities
- single-valued vs. multivalued: one attribute can have multiple values, at a time or over time

Ok for a conceptual data model





# A Data Model Represents

Relationships  
form a concrete  
Business Assertion

**relationships**

among those  
entities and  
**(often implicit)**  
relationships  
among those  
attributes

A relationship called "is the  
placer of" operates on entity  
classes CUSTOMER and  
ORDER and forms the  
following concrete assertion:

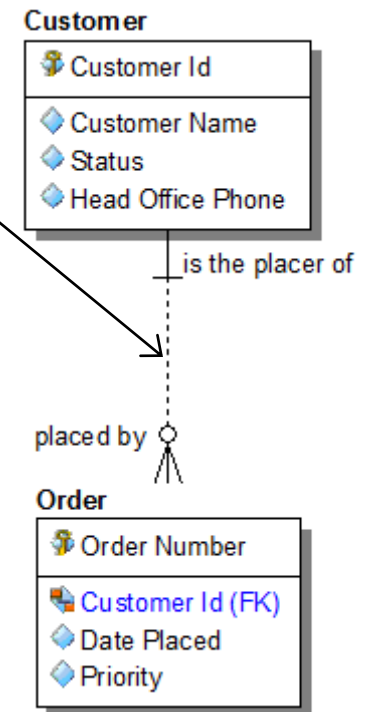
**"Each CUSTOMER  
is the placer of  
zero, one or more  
ORDER(s)"**

Is this true?

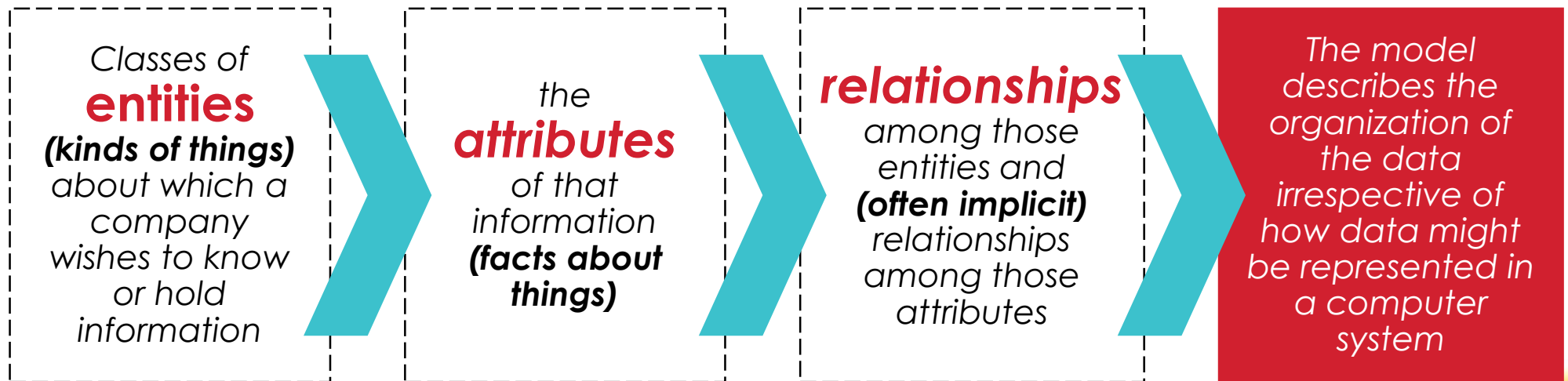
Relationships should be  
named in both directions,  
thus in the other direction  
we have:

**"Each ORDER  
must be placed  
by one and only  
one CUSTOMER"**

Is this true...  
always?



# A Data Model represents



*It's much more than a  
picture!*



# Why Produce A Data Model?

## TOP REASONS\*

1. Capturing Business Requirements
2. Promotes Reuse, Consistency, Quality
3. Bridge Between Business and Technology Personnel
4. Assessing Fit of Package Solutions
5. Identify and Manage Redundant Data
6. Sets Context for Project within the Enterprise
7. Interaction Analysis: Complements Process Model
8. Pictures Communicate Better than Words
9. Avoid Late Discovery of Missed Requirements
10. Critical in Managing Integration Between Systems
11. Pre-cursor to DBMS design / generate DDL

\* DAMA-I Survey

# Why Data Modelling Is Important

Data is an asset  
of your  
corporation

Needs to be  
understood to be  
managed

Don't need to  
look at the detail  
right away (or  
sometimes ever)

An aid to  
understanding

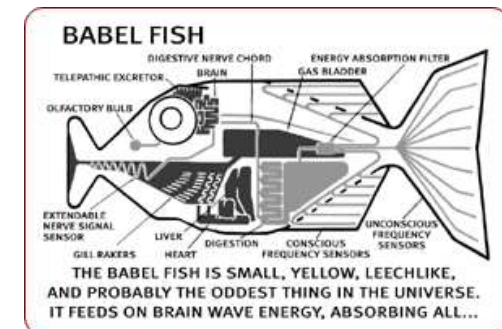
Provides a  
common  
vocabulary

Context and high  
level views will be  
sufficient



Provide only the  
level of detail  
that is necessary  
– fit for purpose

LINGUA  
FRANCA



# Why Data Modelling Is Important

## BUSINESS ARCHITECTURE

*Business  
Objectives & Goals*

—

*Motivations &  
Metrics*

—

*Functions, Roles,  
Departments*

## INFORMATION ARCHITECTURE

*Enterprise Data  
Model*

—

*Conceptual Data  
Models*

—

*Logical Data  
Models*

—

*Physical Data  
Models*

## PROCESS ARCHITECTURE

*Overall Value  
Chain*

—

*High-Level  
Business Processes*

—

*Workflow Models*

## APPLICATION / SYSTEMS ARCHITECTURE

*Systems within  
Scope*

—

*High-Level Mapping*

—

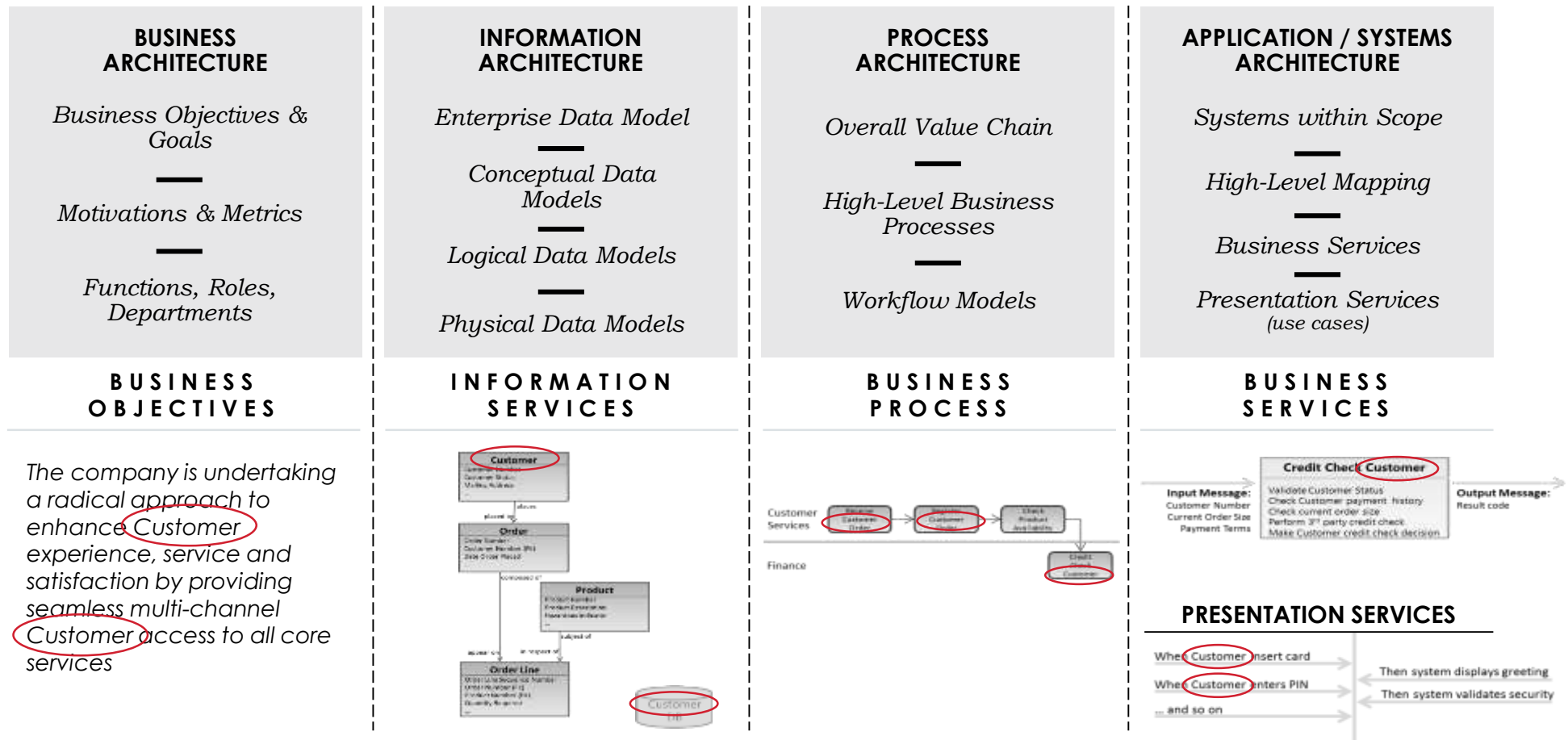
*Business Services*

—

*Presentation  
Services (use cases)*



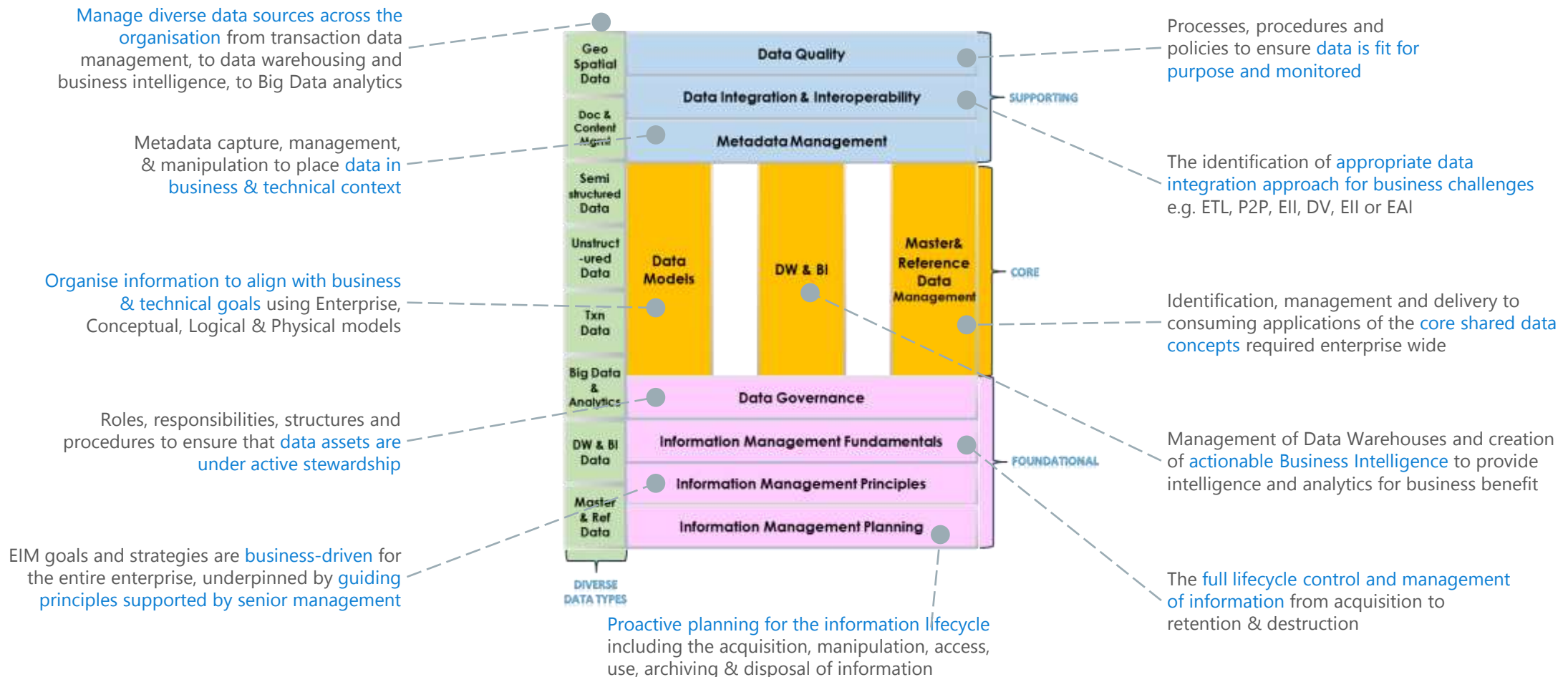
# Why Data Modelling Is Critical



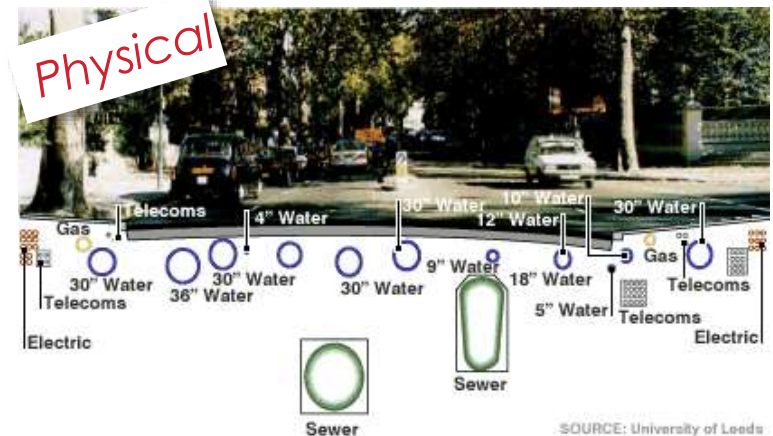
**ALL of the Architecture disciplines use the language (and rules) of the data model**

# Framework for Enterprise Information Management

Development of realistic Information Management strategies to align the desired Information capabilities and services with business motivations and strategies. The information initiatives can be accelerated by use of our Reference Architecture models to understand the capabilities, and typical functional areas for each IM discipline under consideration (such as MDM, DQ, Data Integration etc.). Our Architecture Reference models contain the typical areas of functionality & capabilities observed in each IM discipline. Our EIM framework has capability & maturity models for each of the IM disciplines together with the typical processes and activities observed in mature organisational services for each.



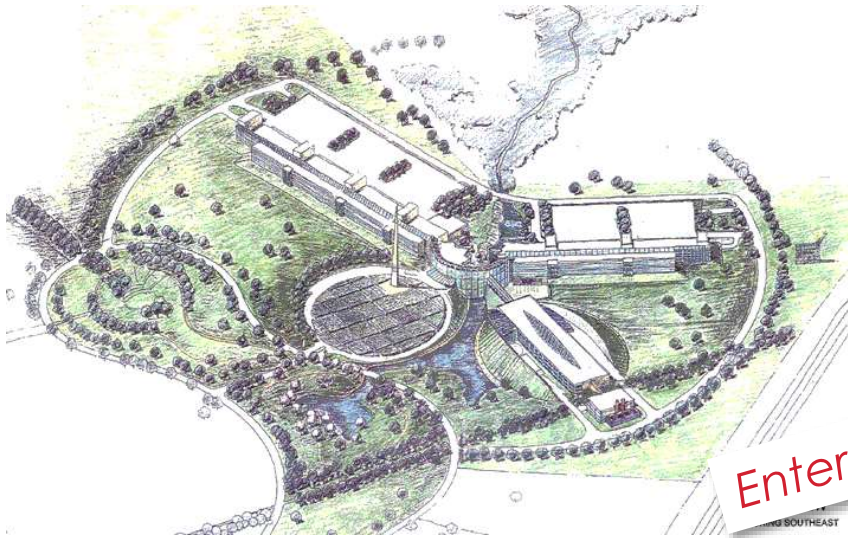
# We All Use Models



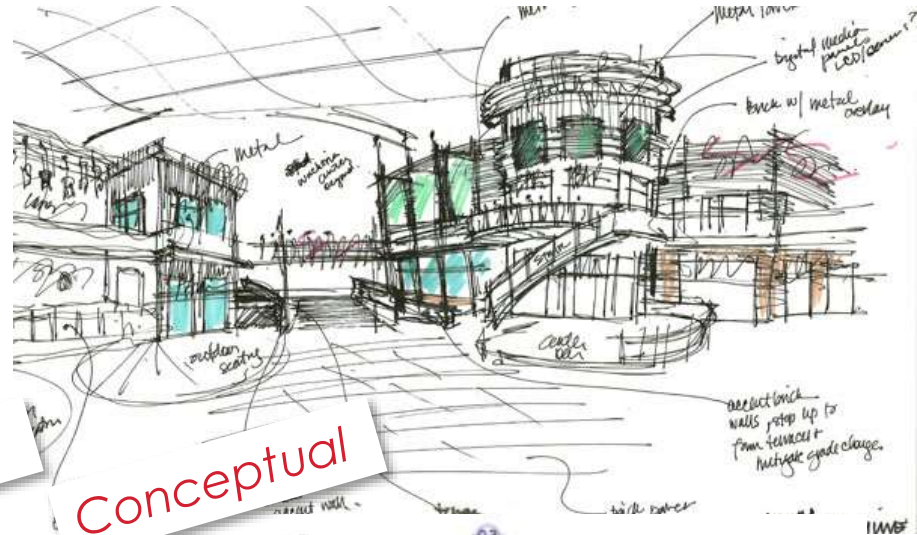
SOURCE: University of Leeds



# We All Use Models



Enterprise



Conceptual

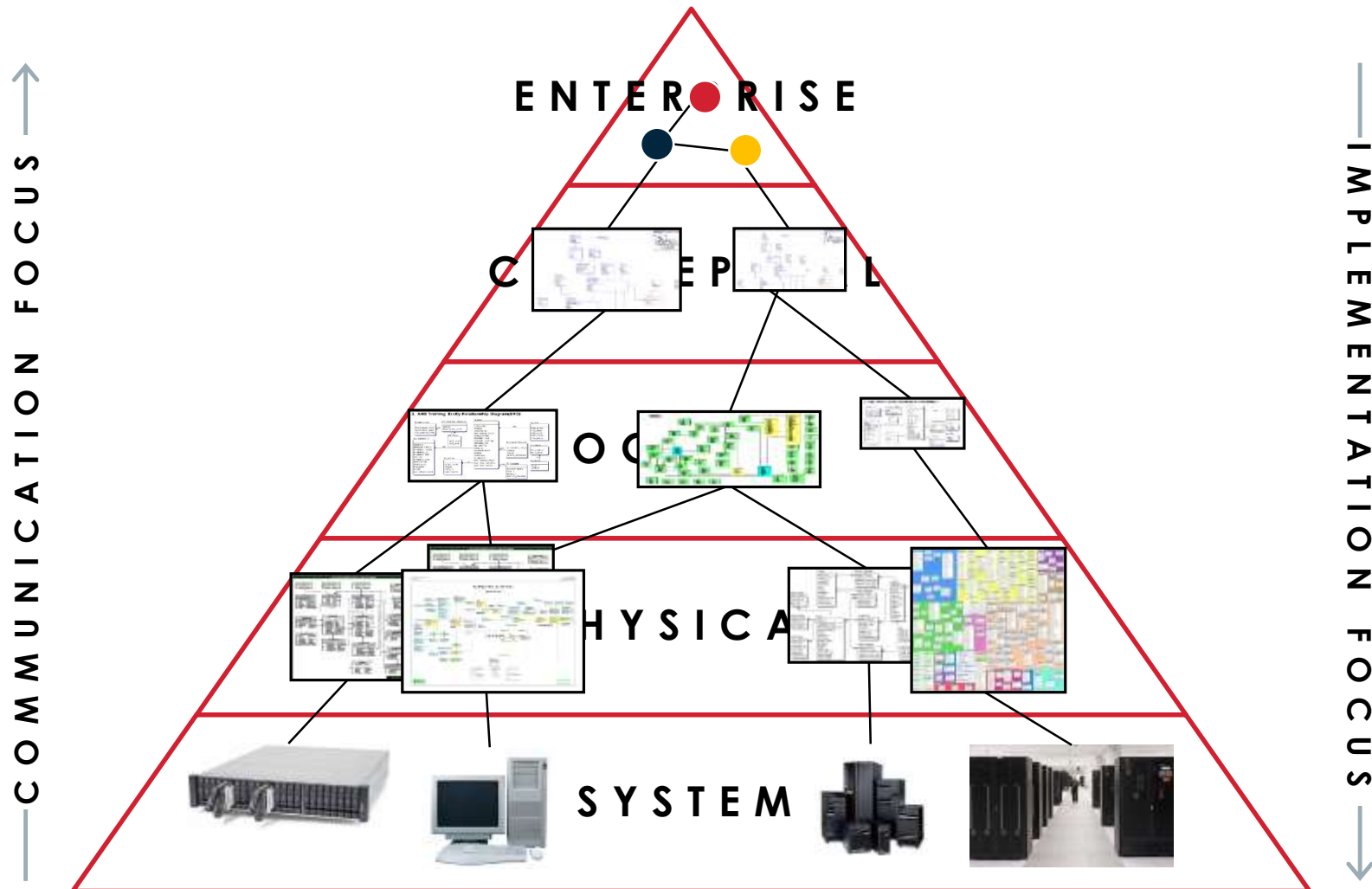


Logical



Physical

# Data Model Levels





# Levels of Data Models

## Enterprise Data Model

Documents the very high level business data objects and definitions. Enterprise wide scope to provide a strategic view of Enterprise data.

## Conceptual Data Model (Subject area)

The business key, attributes and definitions of business data objects. Also shows the relationship between business data objects. Broader scope than LDM and may cover a subject area (also known as subject area data model).

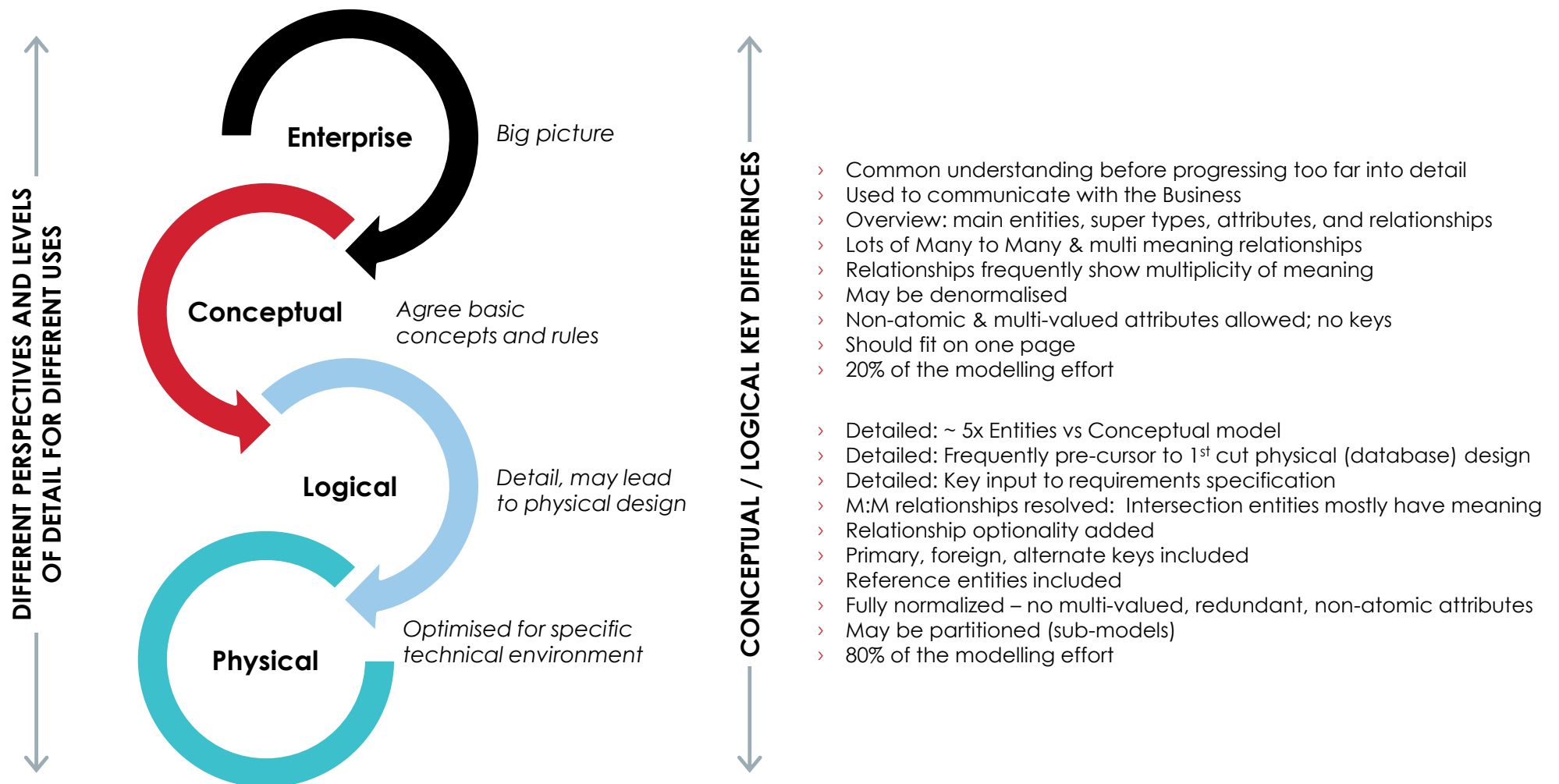
## Logical Data Model (Application)

Documents the business key, attributes and definitions of business data objects. It also shows the relationship between business data objects. Frequently is within the scope of a defined project.

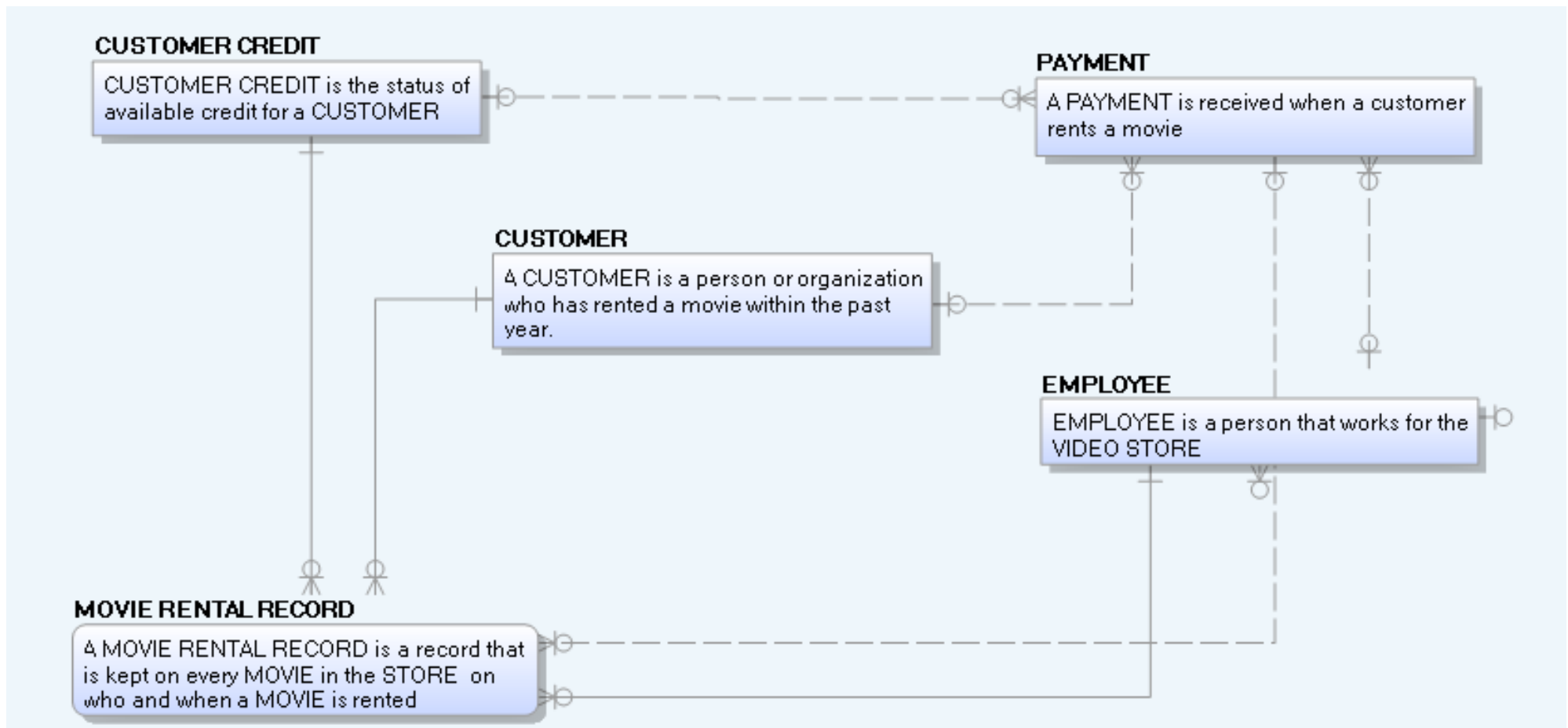
## Physical Data Model

Technical design eg tables, columns, keys, foreign keys, and other constraints to be implemented in the database or XSD. May be generated from a logical data model. This model is within the scope of a defined project.

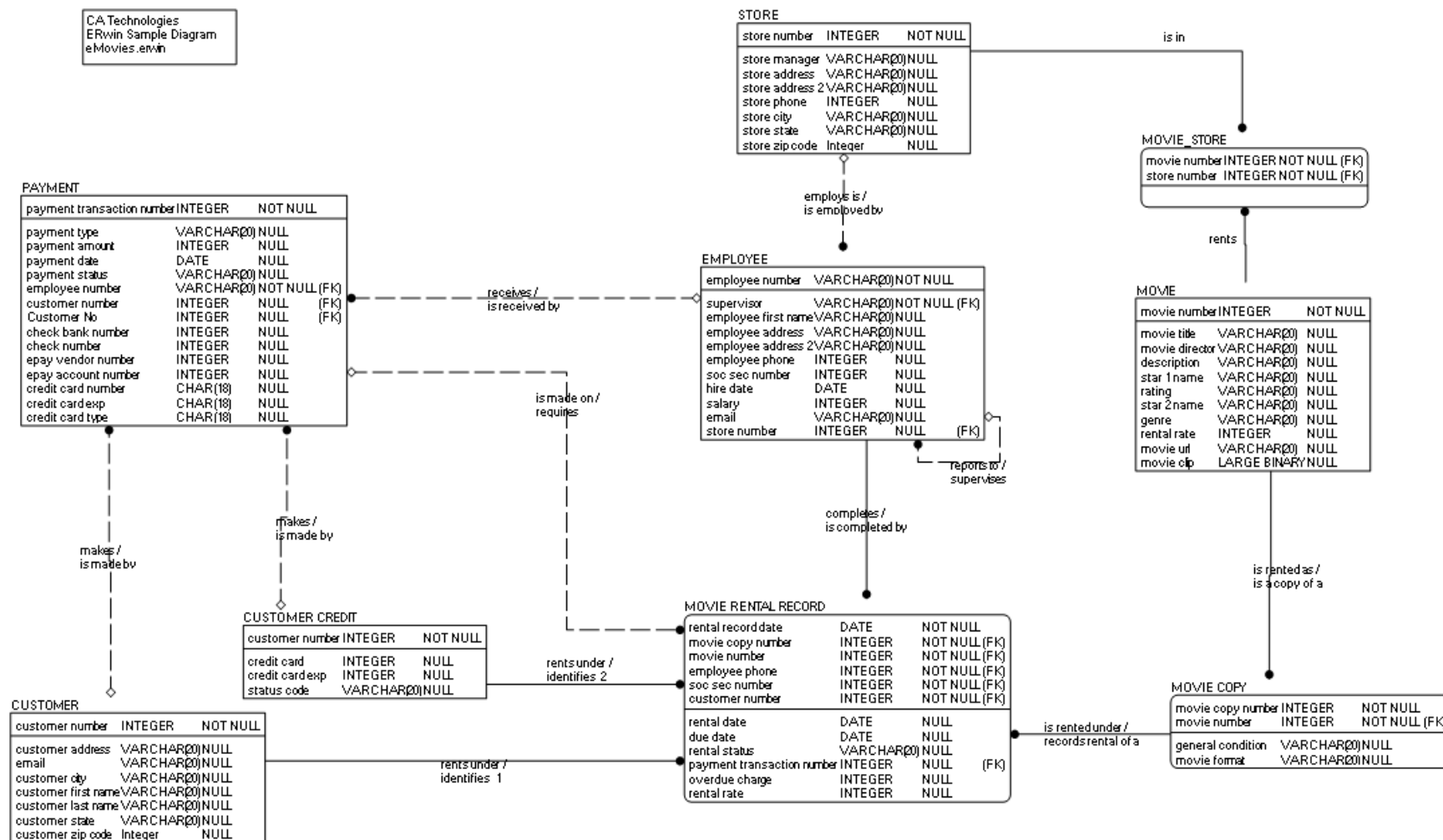
# Enterprise vs. Conceptual vs. Logical



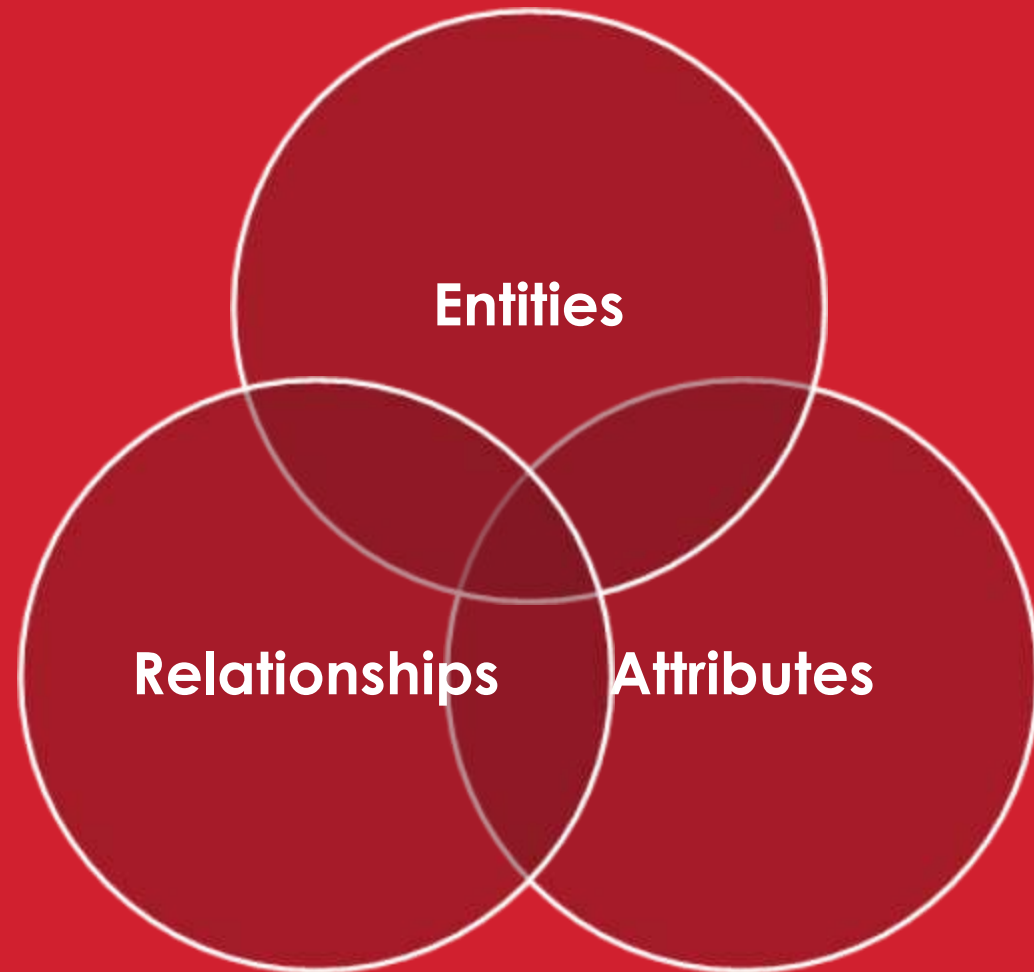
# Different Data Models For Different Audiences: *Business*



# Different Data Models For Different Audiences: *Technical*

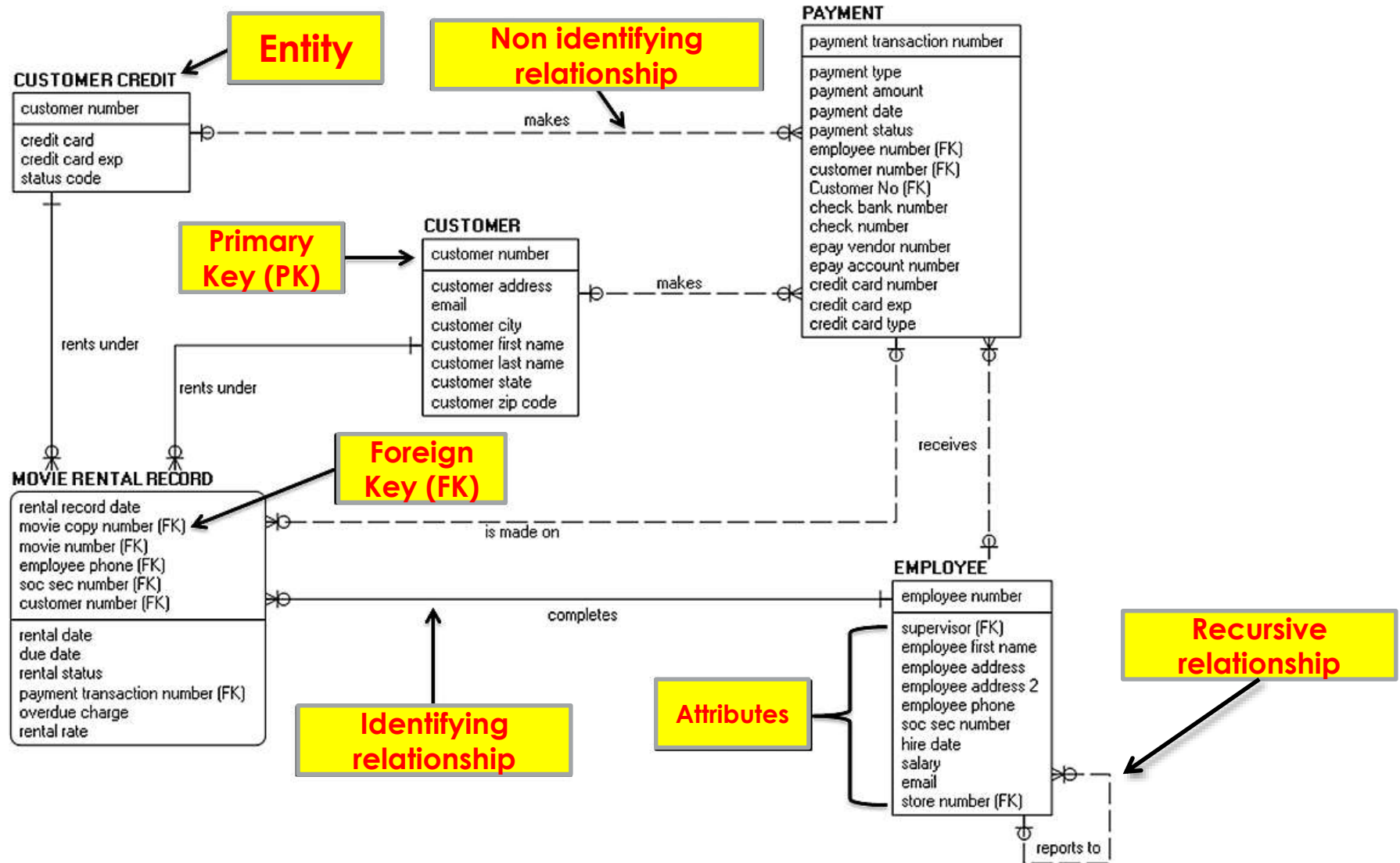


# Logical Data Modelling Components





# Logical Data Model Components



# Entities

**A THING OF SIGNIFICANCE TO THE BUSINESS ABOUT  
WHICH INFORMATION NEEDS TO BE KNOWN OR HELD**

**WHO**

*Person, Employee, Vendor, Customer,  
Department, Organisation, ...*

**WHAT**

*Product, Service, Raw Material, Training  
Course, Flight, Room, ...*

**WHEN**

*Time, Day, Date, Calendar, Reporting Period,  
Fiscal Period, ...*

**WHERE**

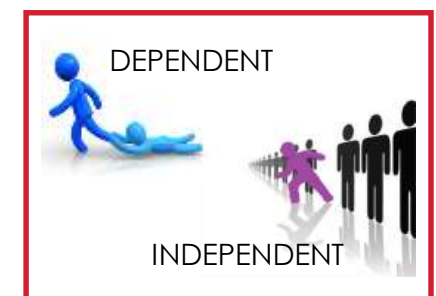
*Geographic location, Delivery address,  
Storage Depot, Airport, ...*

**WHY**

*Order, Complaint, Inquiry, Transaction, ...*

**HOW**

*Invoice, Policy, Contract, Agreement,  
Document, Account, ...*



# Entity Naming Best Practice

Entity names must be unique  
(e.g. Order)

*Can be accomplished by placing qualifiers or quantifiers in front of the entity name.*

Entity names must have  
business meaning

*Use a noun alone wherever possible (e.g. **Contract**)*

Use adjective + noun or,  
adjective + adjective + noun  
to clarify meaning

*e.g. **Lease Contract** or **Back Order***

Entity names should be in  
Title Case format

*e.g. “**Product Group**” not “**PRODUCT GROUP**”*

Entity names should be  
singular

*The entity is defined in terms of a single occurrence.  
(e.g. “**Product**” not “**Products**”)*

Acronyms or abbreviations  
should not be used

*Do not use “\_” within **Entity** names, this really is only  
needed for implementation artefacts (e.g. **Tables**)*

# Entity Definition Best Practice

## CLARITY

- › Can the term be understood by reading the definition?
- › Is it unambiguous?
- › Avoid jargon & stating the obvious

## COMPLETENESS

- › Appropriate level of detail – not too generic – not too specific
- › Goldilocks principle
- › Contains all necessary components without omission (e.g. derivations, UOM)

## ACCURACY

- › A subject matter expert would agree
- › Relevant to the state of the entity: i.e. the stages an entity may go through over time



# Common Errors With Entities

*A few common mistakes are regularly encountered when creating business focused data models:*

- › Creating an Entity that is really a report or a screen or form
- › Failing to clarify if the entity deals with types (or categories) vs. specific instances of things
- › Identifying an entity that exists in the real world, but whose instances can't be uniquely identified e.g. "Billboard observer"
- › Identifying entities that are too imprecise and / or the name doesn't imply a single instance e.g. "Weather"
- › Crossing to the "techno side" and introducing implementation specific constructs





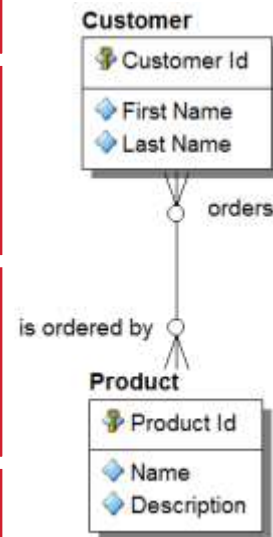
# Dependent

Does not depend on any other entity for its existence.

Can be identified without reference to another entity on the model

You don't need to know anything about a Customer to identify a Product

You don't need to know anything about a Product to identify a Customer



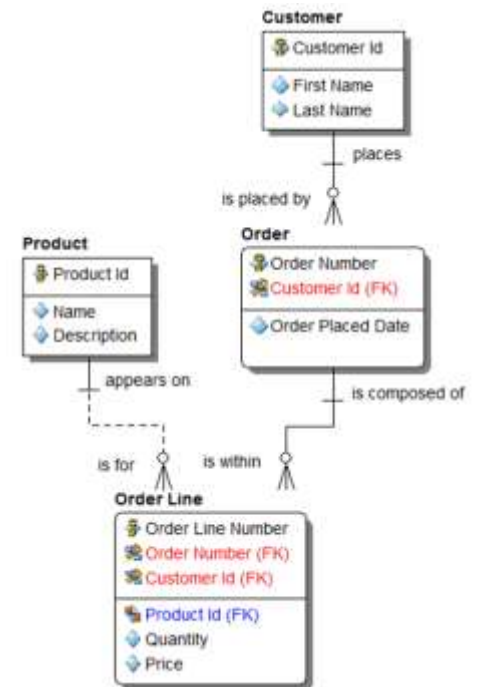
# Independent

Depends upon one (or more) other entities

Can depend upon Independent or Dependent entities

3 types:

1. Attributive (depends only on its parent)
2. Associative (depends upon two or more entities)
3. Category (AKA supertype)



Why is it important to know whether an entity is independent or dependent?



# Attributes & Keys

**KEYS (SUMMARY):**  
**AN ENTITY CAN CONTAIN FOUR TYPES OF KEYS**

## PRIMARY KEY (PK)

- › Identifies an entity/table for the system.
- › May be "natural" or "surrogate"
- › Could be composite

## ALTERNATE KEY (AK)

- › Another way to identify an entity/table.
- › May be surrogate

## FOREIGN KEY (FK)

- › Identifies a relation among entities/tables

## INVERTED KEY (IK) OR INVERSION ENTITY [IE]

- › Improves access to table information (physical)

*Surrogate is often a "made up" key*

Composite

is made up

of several

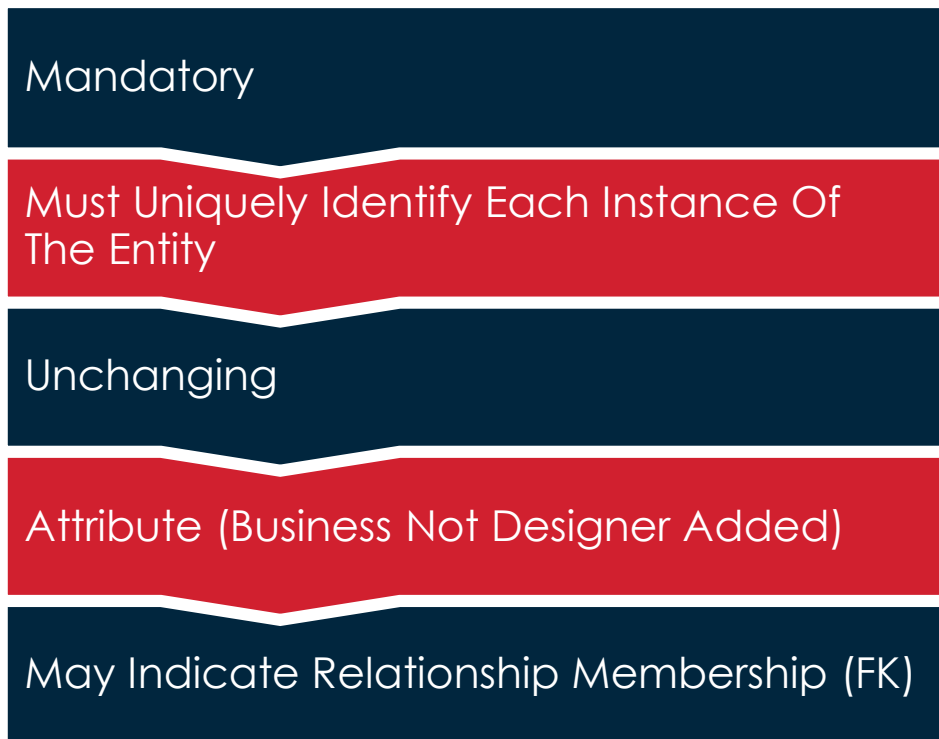
items to

form the key

*Entity Keys*

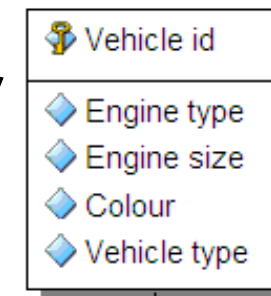


# Primary Key



*Attribute(s) that make up a PK are represented in modelling tools separately from the rest of the attributes by a line*

## Vehicle



Can I use  
Registration  
Number as the  
PK?



# Primary Keys

- › What attributes might uniquely identify an entity? Let's use **Customer** as an example.
- › What might uniquely identify an individual customer?

Is **Last Name + First Name** enough?

— Could there be 2 customers named *John Smith*? → Probably

Is **Last Name + First Name + Date of Birth** enough?

— Could there be 2 customers named *John Smith* born on *1 June, 1963*? → Less Likely, but Possible

Is **Last Name + First Name + Date of Birth + Address** enough?

— Could there be 2 customers named *John Smith* born on *1 June, 1963* living at *1 Earl's Court, London, UK*? → Even Less Likely. Possible, but how many attributes do we want to use?

# Keys:

## Natural vs. Surrogate

- › The “Customer” example keys we just identified would be classified as *natural* keys.
- › Natural keys are based on business rules and logic that determine how an individual instance can be uniquely identified.
- › As we’ve seen, natural keys can become unwieldy, requiring a number of attributes, which makes queries difficult.
- › Also, extreme care is needed as components of natural keys could change

— **Surrogate keys** are often used instead, which are system-generated unique identifiers. e.g. Customer ID, Product ID, etc.

— While surrogate keys are more efficient, important business rules are lost when they are used. ***It’s a balancing act.***



# Alternate Key

**ALTERNATE KEY (AK):**  
**ANOTHER WAY OF IDENTIFYING THE ENTITY**

*Are these good  
AK's for all  
circumstances?*



## Person

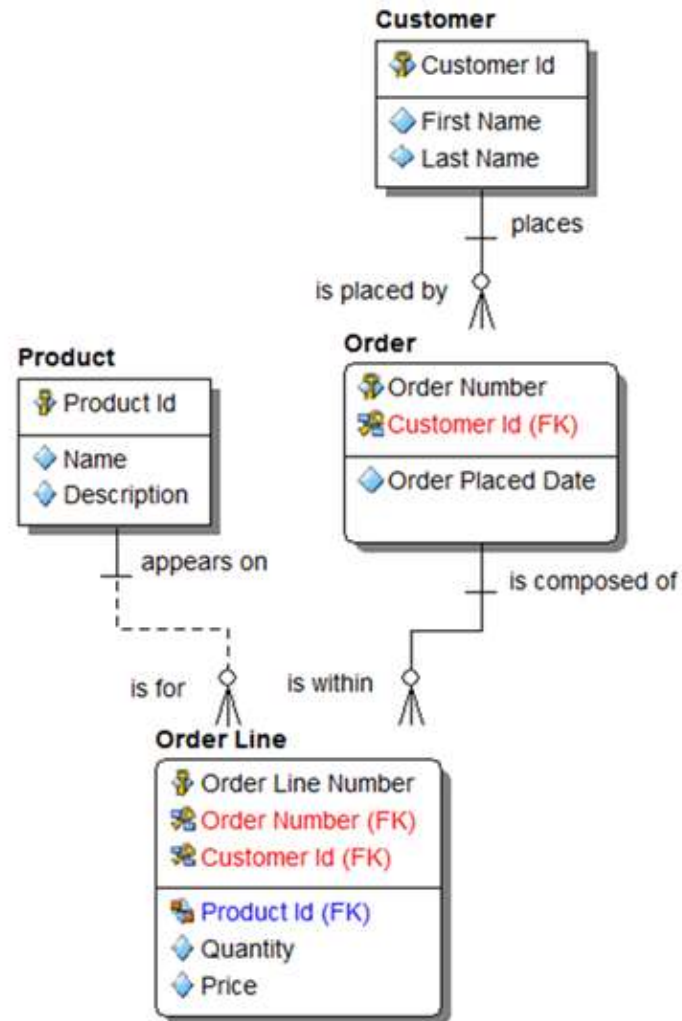
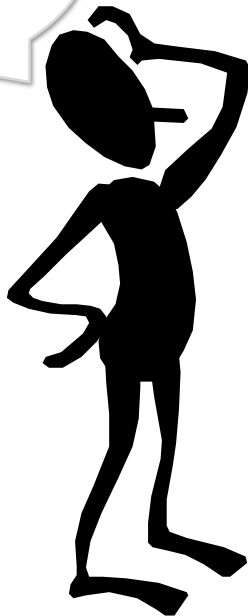
🔑 Person id		
◆	Passport number	(AK2:1)
◆	Social security number	(AK1:1)
◆	Name	
◆	Date of birth	
◆	Nationality	



# Foreign Key

**FOREIGN KEY (FK):**  
**POINTS BACK TO THE PARENT(S)**

Can a FK be  
part of the  
Primary Key?





# Inverted Key

**INVERTED KEY (IK):  
PHYSICAL IMPLEMENTATION ONLY**

**TABLE**

Primary Key	Car	Eye Colour	Nationality	Salary
1	Audi	Blue	UK	100000
2	BMW	Green	US	95000
3	BMW	Brown	FR	85000
4	Bentley	Blue	UK	250000
5	Audi	Brown	US	60000
6	Ford	Blue	FR	50000
7	Ford	Brown	UK	45000
8	Audi	Brown	US	55000
9	BMW	Blue	UK	65000

## CAR IK

Audi (1, 5, 8)  
Bentley (4)  
BMW (2, 3, 9)  
Ford (6, 7)

## EYE COLOUR IK

Blue (1, 4, 6, 9)  
Brown (3, 5, 7, 8)  
Green (2)

## NATIONALITY IK

FR (3, 6)  
UK (1, 4, 7, 9)  
US (2, 5, 8)

## IK IMPLEMENTED AS BITMAP INDEX

	1	2	3	4	5	6	7	8	9
Car = Audi	1				1			1	
Car = Bentley				1					
Car = BMW		1	1						1
Car = Ford						1	1		
Eyes = Blue	1			1		1			1
Eyes = Brown			1		1		1	1	
Eyes = Green		1							
Nationality = FR			1			1			
Nationality = UK	1			1			1		1
Nationality = US		1			1			1	

# Attribute Properties & Domains

Name

- Logical
- Physical

Unique

Part of primary key

Mandatory

Datatype

Domain

Validation Rules

Default Value

Nullable (Y/N)

Definition

Notes

*Main Attribute Properties*



THE COMPLETE SET OF VALID  
VALUES A DATA ELEMENT MAY  
CONTAIN (EG DROP DOWN)

Characteristics of  
an attribute

DOMAINS CAN BE CASCADED  
TO OTHER ATTRIBUTES

E.g.

**Person Id**

*"The unique identifier of a Person in Acme corporation. This is a unique integer length 13, with digit 13 being a checksum"*

**Employee Id:**

Domain name = Person Id

**Business Sponsor Id:**

Domain name = Person Id

**Cost Centre Owner:**

Domain name = Person Id

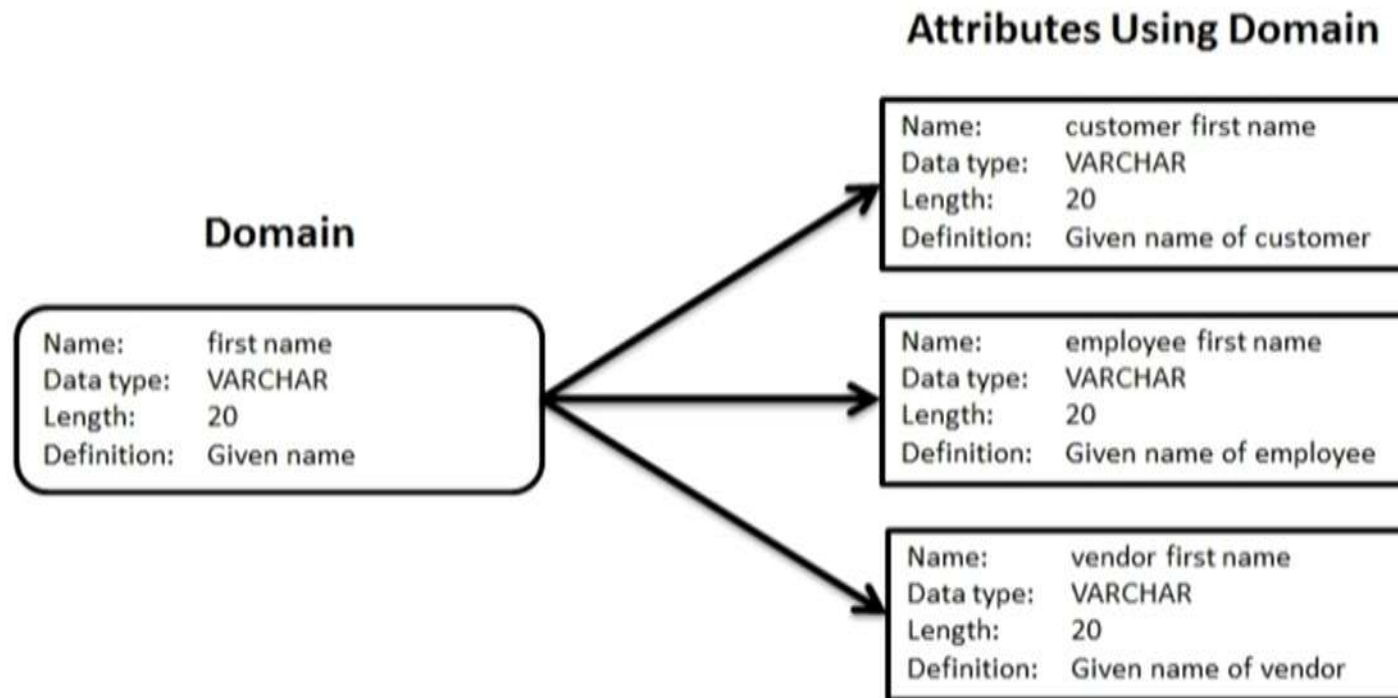
**Customer Id:**

Domain name = Person Id

**Customer Name:**

Domain name = Person Name

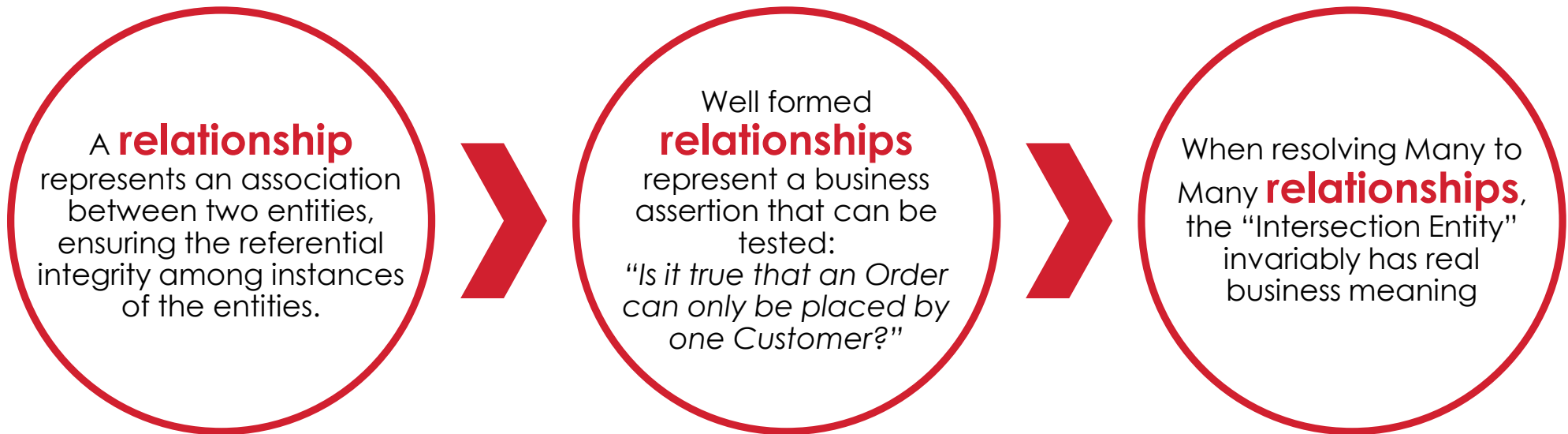
# Domain Inheritance



# Relationships

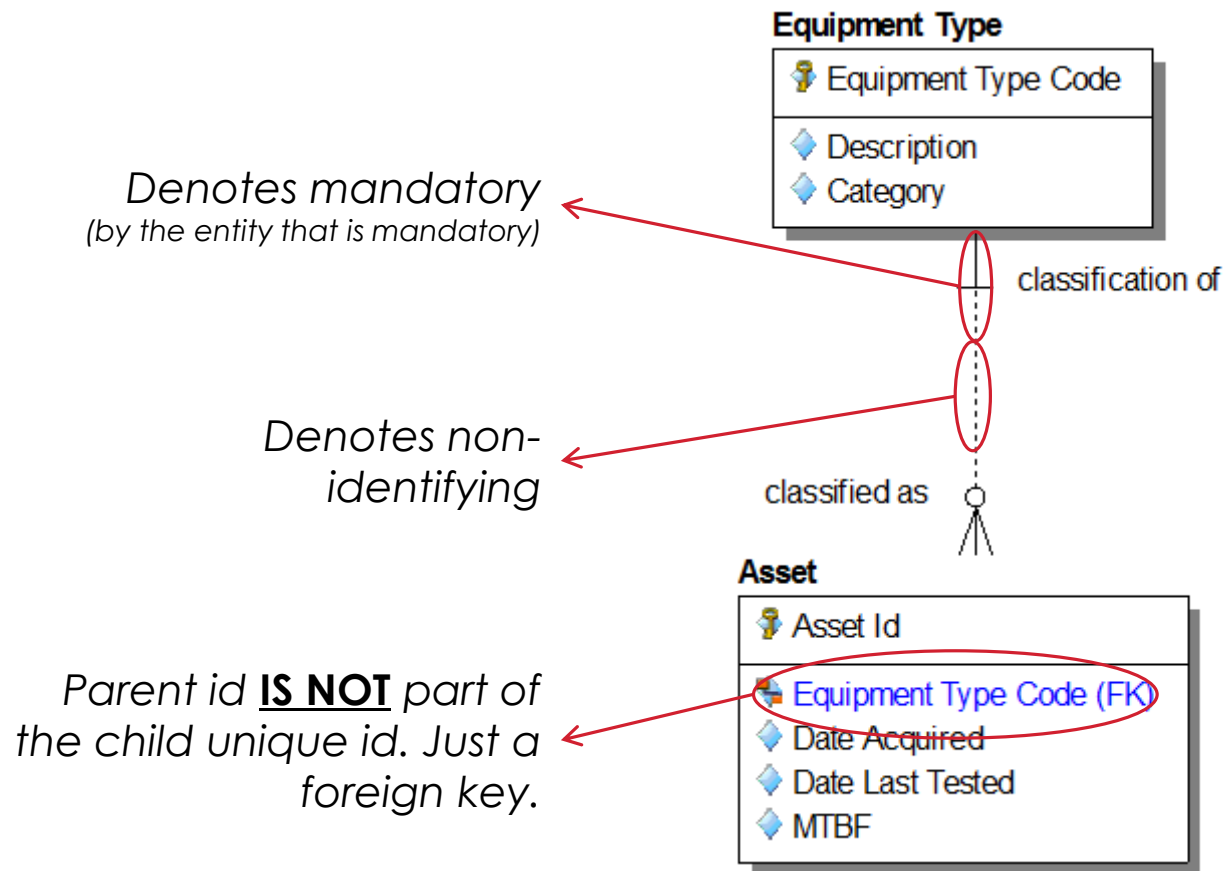


# Relationships Between Entities



# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: MANDATORY, NON-IDENTIFYING



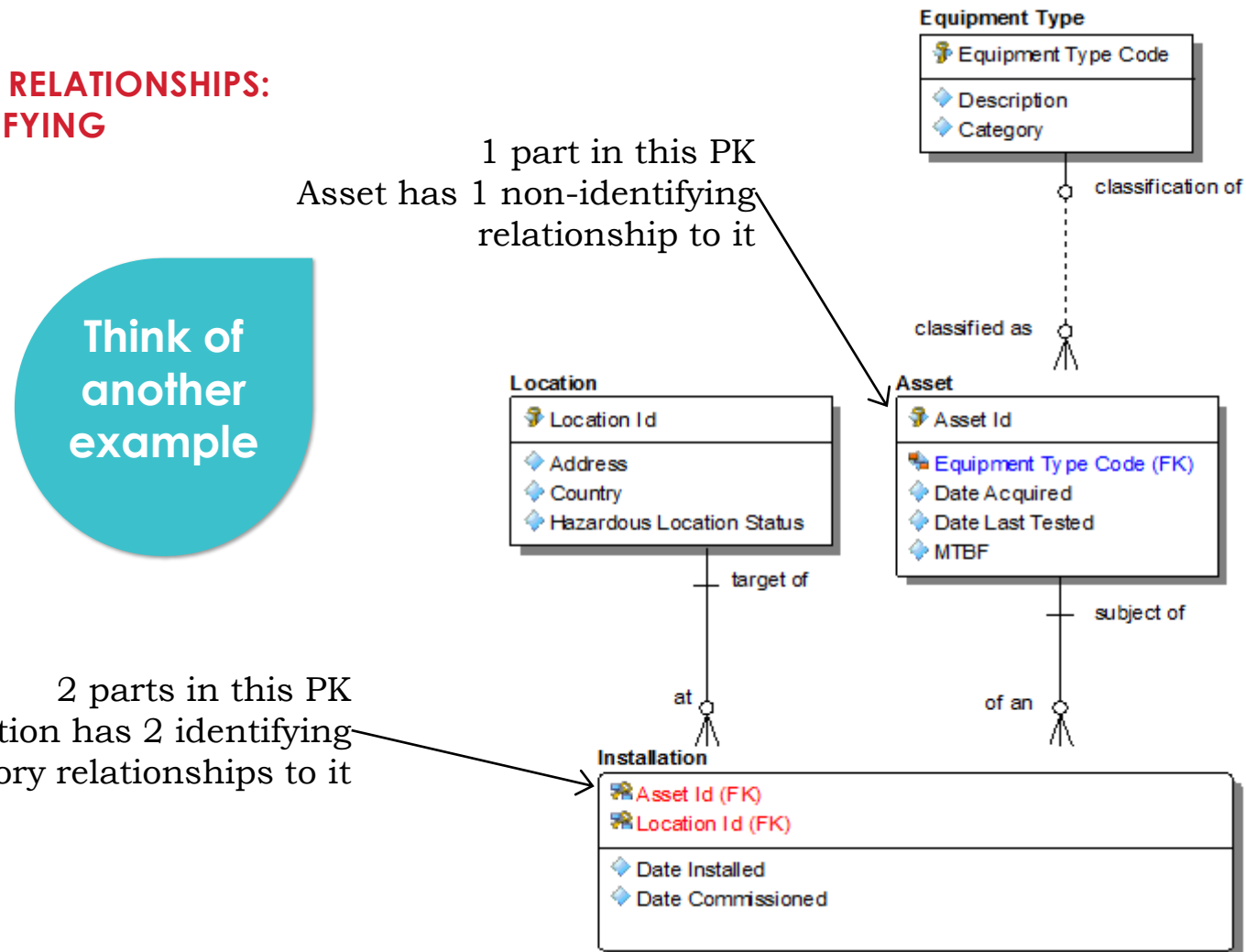
What does this  
tell us about  
Asset Id?





# Relationship Types

TYPES OF “ONE TO MANY” RELATIONSHIPS:  
MANDATORY, NON-IDENTIFYING



# Relationship Types

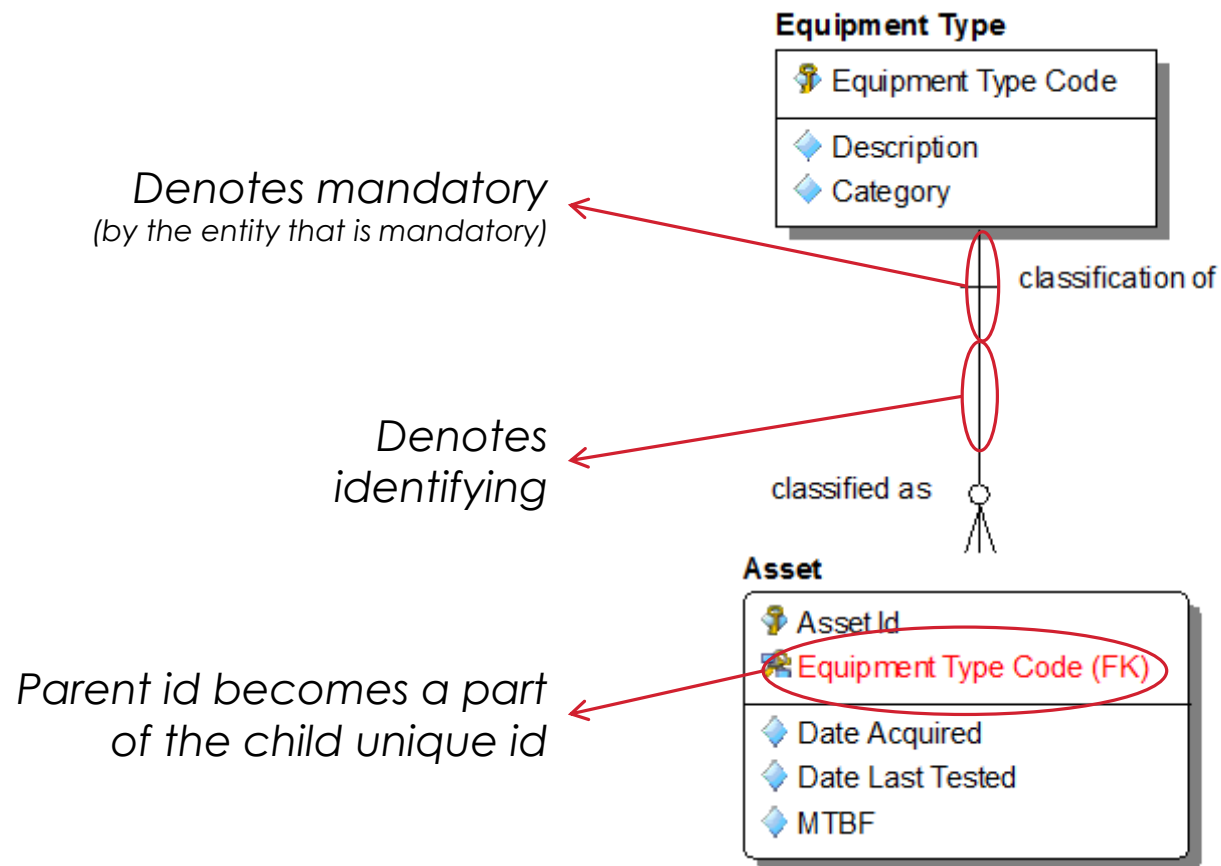
TYPES OF “ONE TO MANY” RELATIONSHIPS:  
MANDATORY, NON-IDENTIFYING



This tells us that the Order  
Number **IS** unique

# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: MANDATORY, IDENTIFYING



What does this tell us about Asset Id?

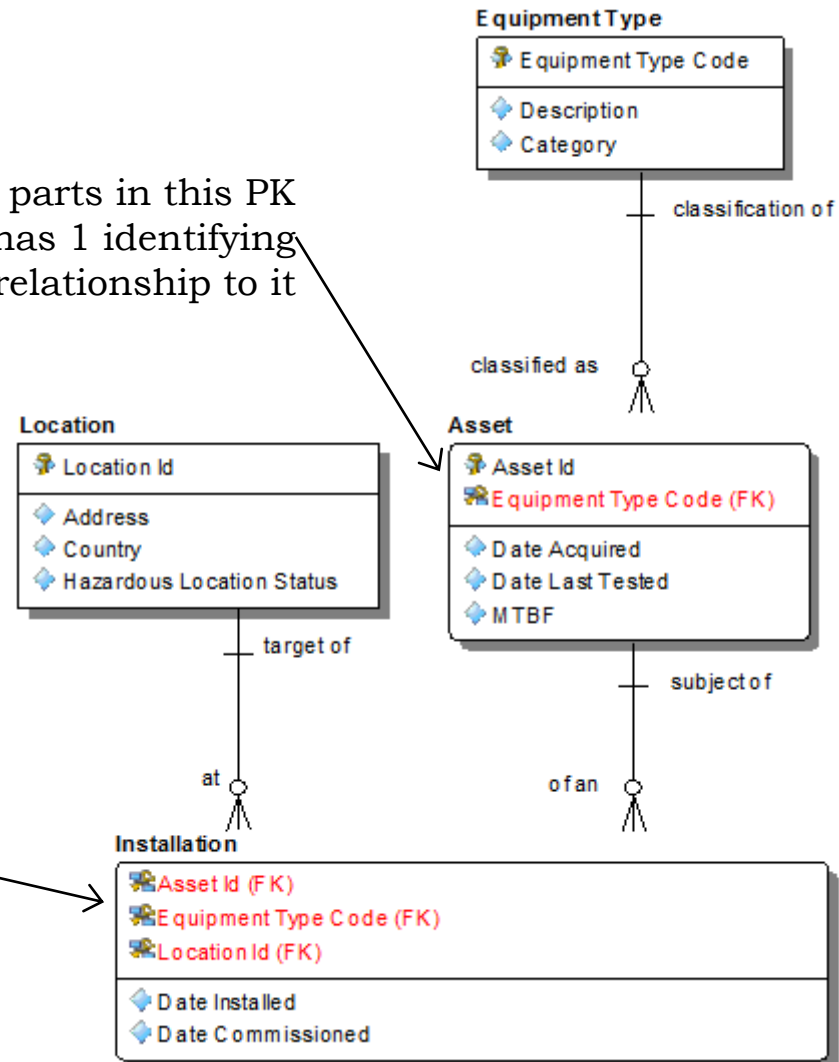
Think of another example

# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: MANDATORY, IDENTIFYING

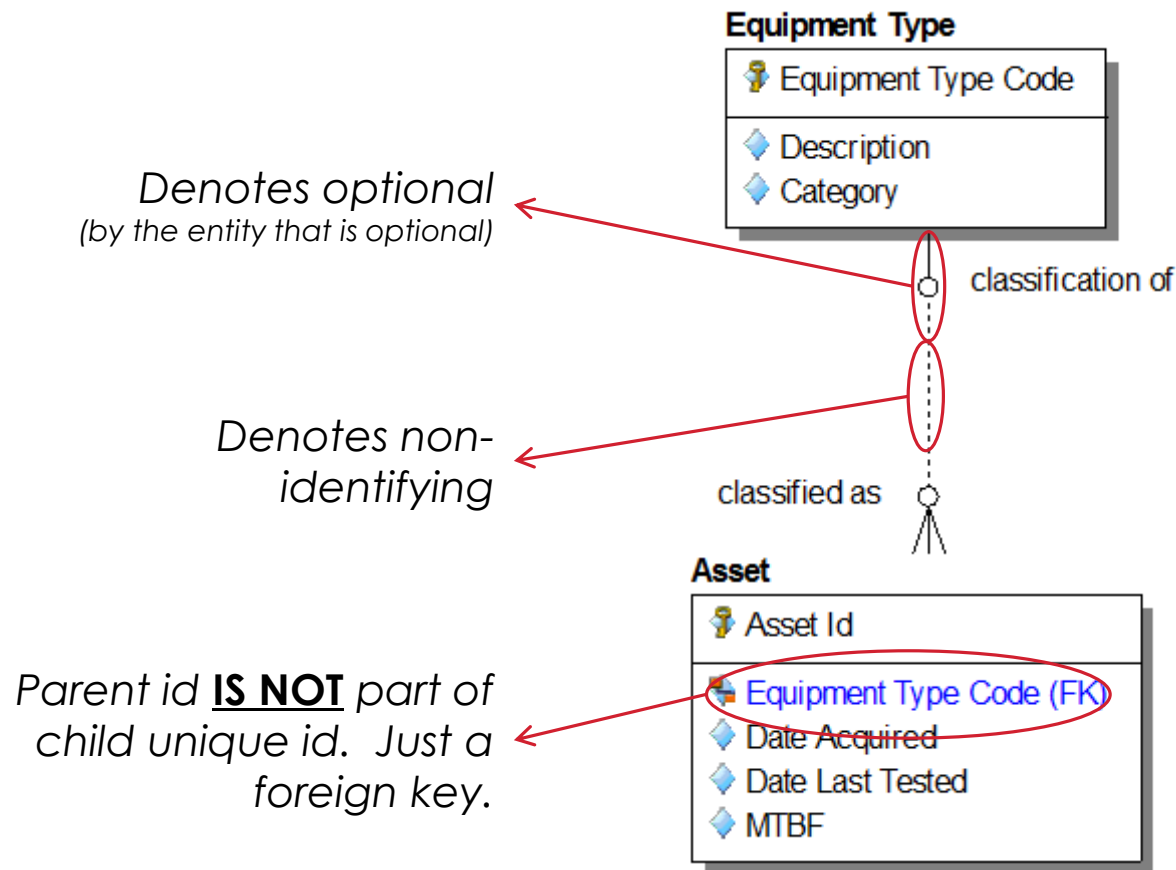
Now there's 2 parts in this PK  
Asset has 1 identifying  
mandatory relationship to it

Installation has 2  
*identifying mandatory*  
relationships to it



# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: OPTIONAL, NON-IDENTIFYING



What does this  
tell us about  
Asset Id?

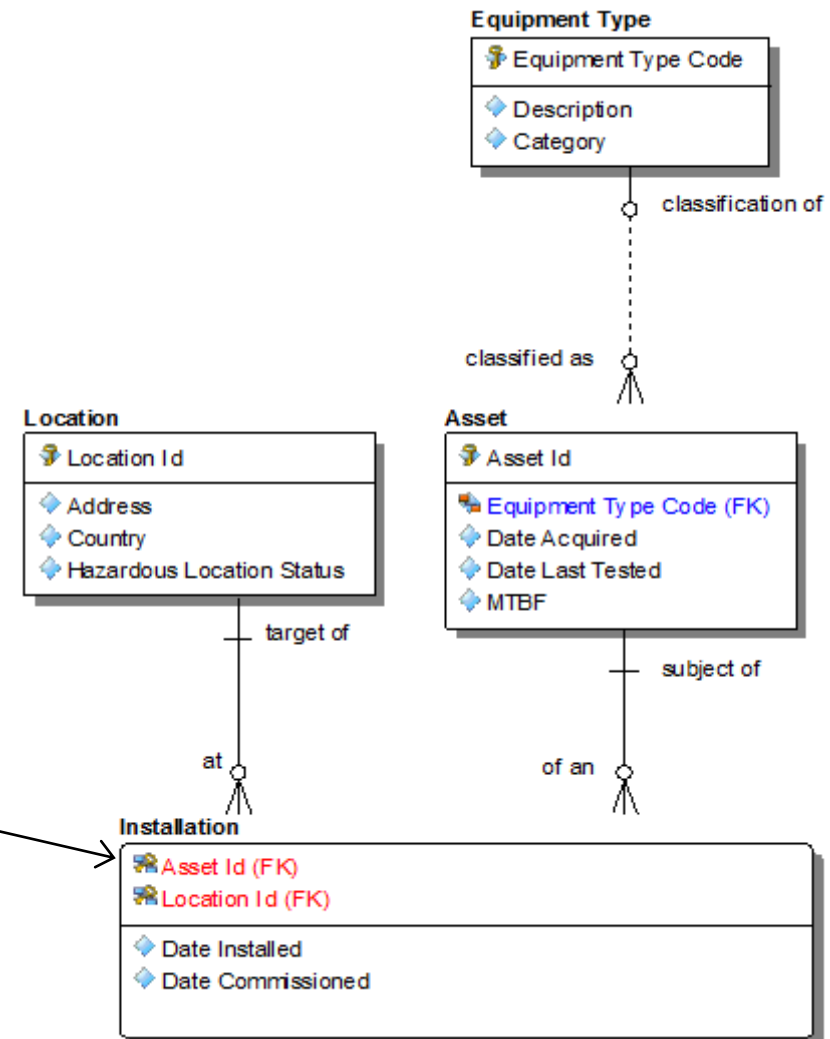


# Relationship Types

TYPES OF “ONE TO MANY” RELATIONSHIPS:  
OPTIONAL, NON-IDENTIFYING

Think of  
another  
example

Installation still has 2 identifying mandatory relationships to it but the primary key has changed as it now doesn't include Equipment Type Code





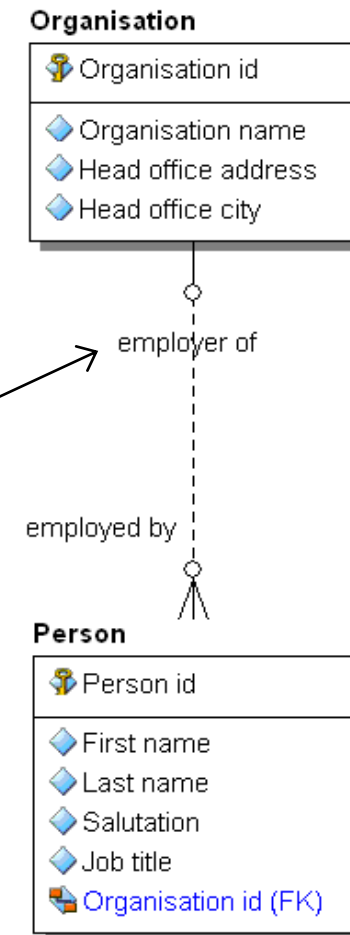
# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: OPTIONAL, NON-IDENTIFYING

This tells us that we don't need to know the Organisation to find a Person. A Person does **NOT** have to be “**employed by**” an Organisation.

Back to  
our Asset  
example

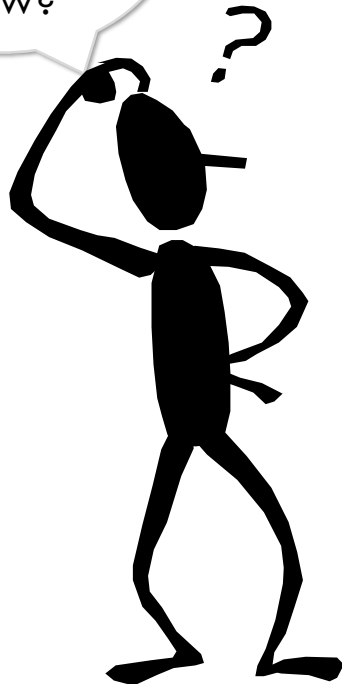
**Organisation** to **Person**  
relationship is optional,  
non-identifying



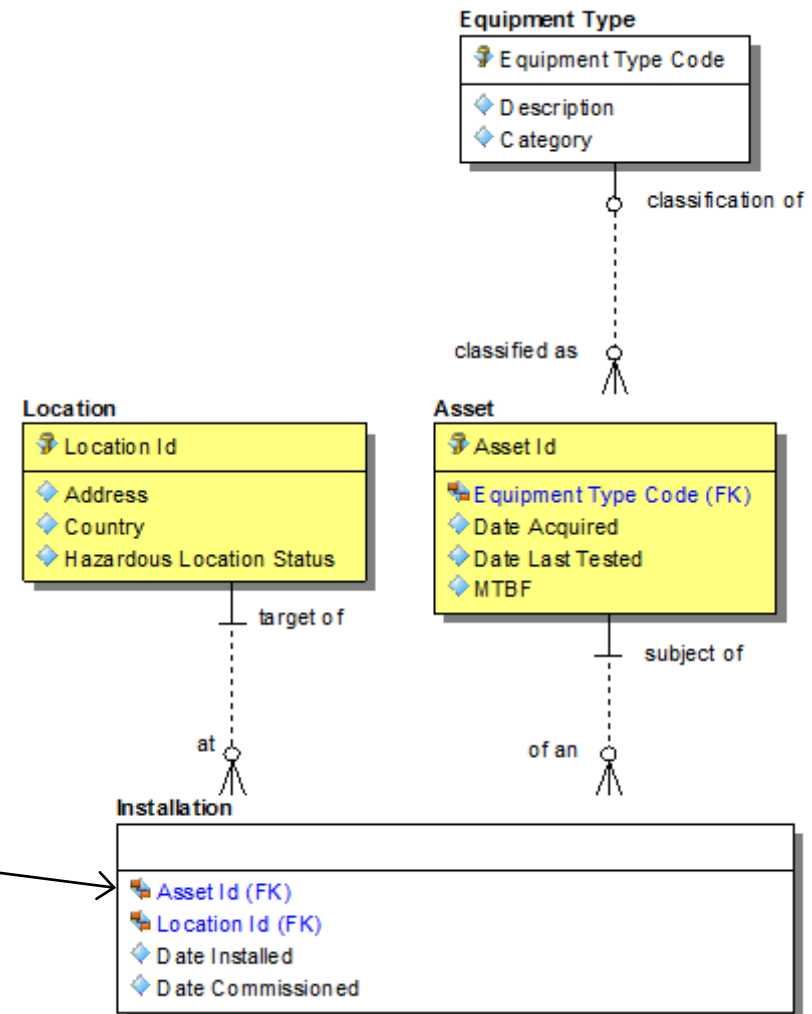
# Relationship Types

## TYPES OF “ONE TO MANY” RELATIONSHIPS: OPTIONAL, NON-IDENTIFYING

So what  
have we  
got to do  
now?



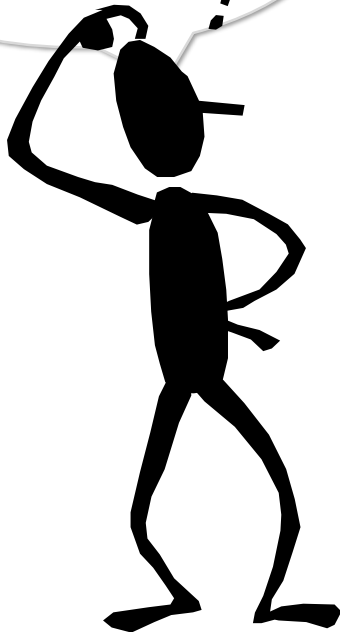
See what happens to  
Installation if its 2  
relationships become  
**non-identifying**.



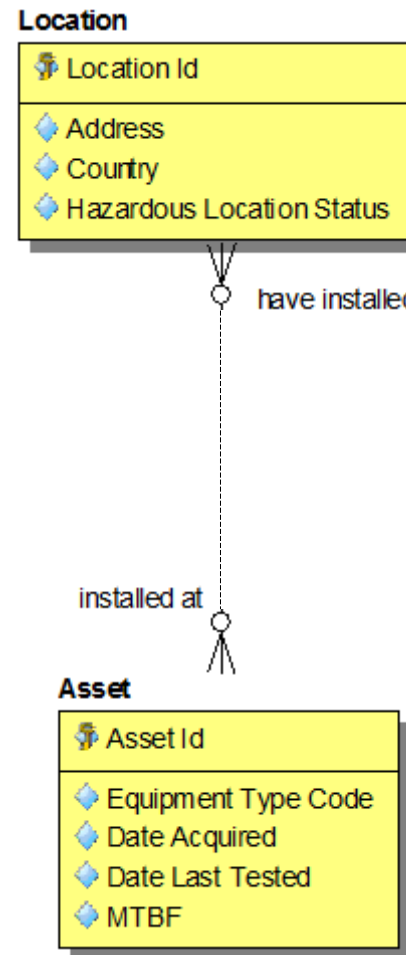
# Relationship Types

TYPES OF “ONE TO MANY” RELATIONSHIPS:  
MANY TO MANY (AKA NON SPECIFIC)

Why can't a many  
to many relationship  
be an “identifying”  
relationship?



Think of  
another  
example



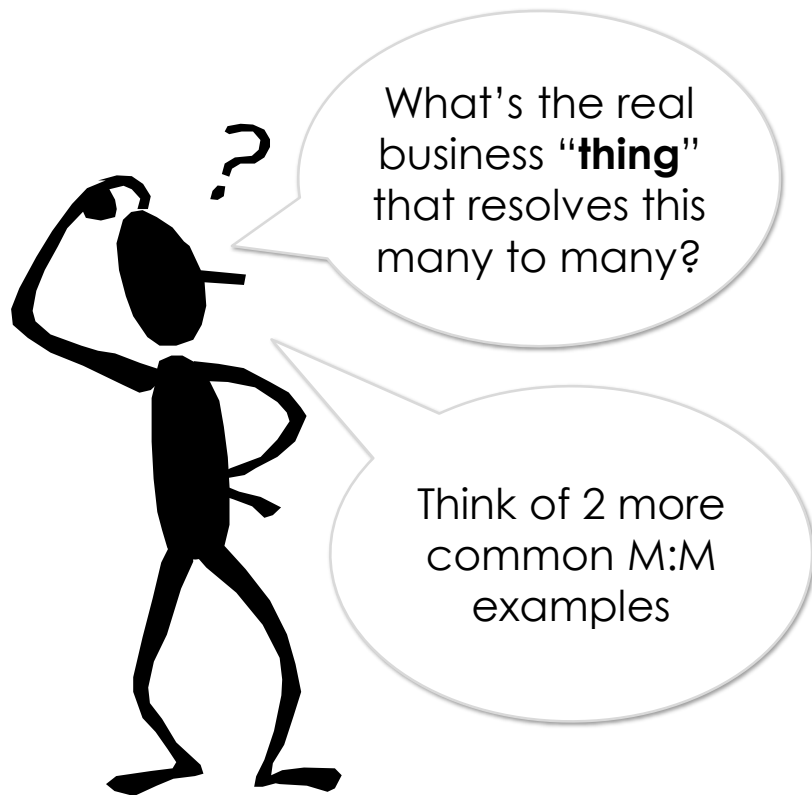
## Chris's law

99% of M:M relationships  
represent a real business  
concept that is the  
intersection entity

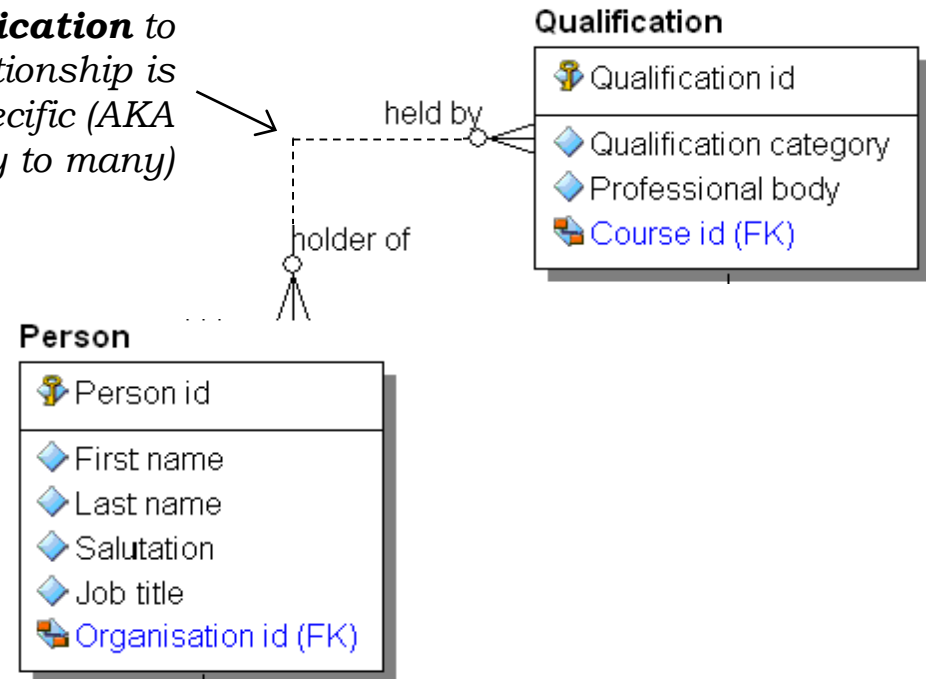


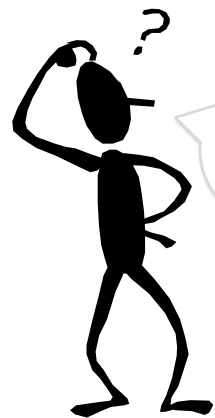
# Exercise 3: Relationships

TYPES OF "ONE TO MANY" RELATIONSHIPS:  
MANY TO MANY (AKA NON SPECIFIC)

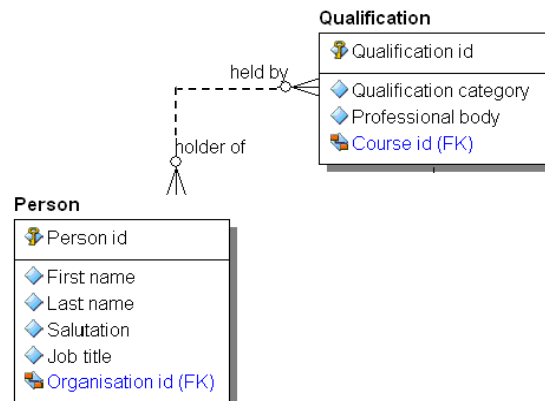


**Qualification** to **Person** relationship is non-specific (AKA many to many)

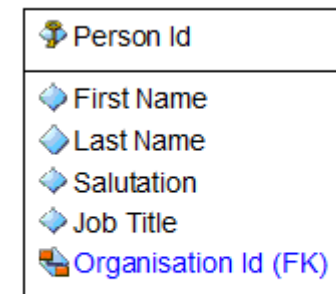




What's the real business **"thing"** that resolves this many to many?



### Person



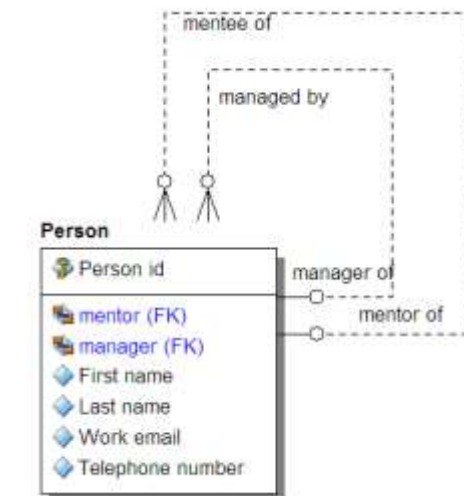
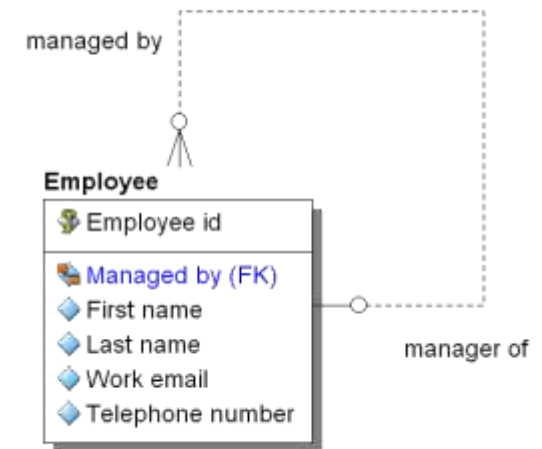
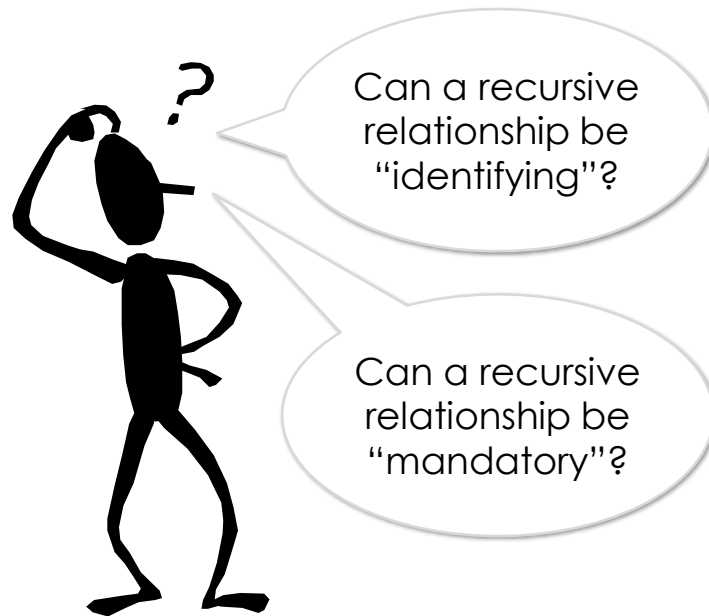
### Qualification



### Award



# Recursive Relationships



A recursive relationship occurs when there is a relationship **between an entity and itself**.



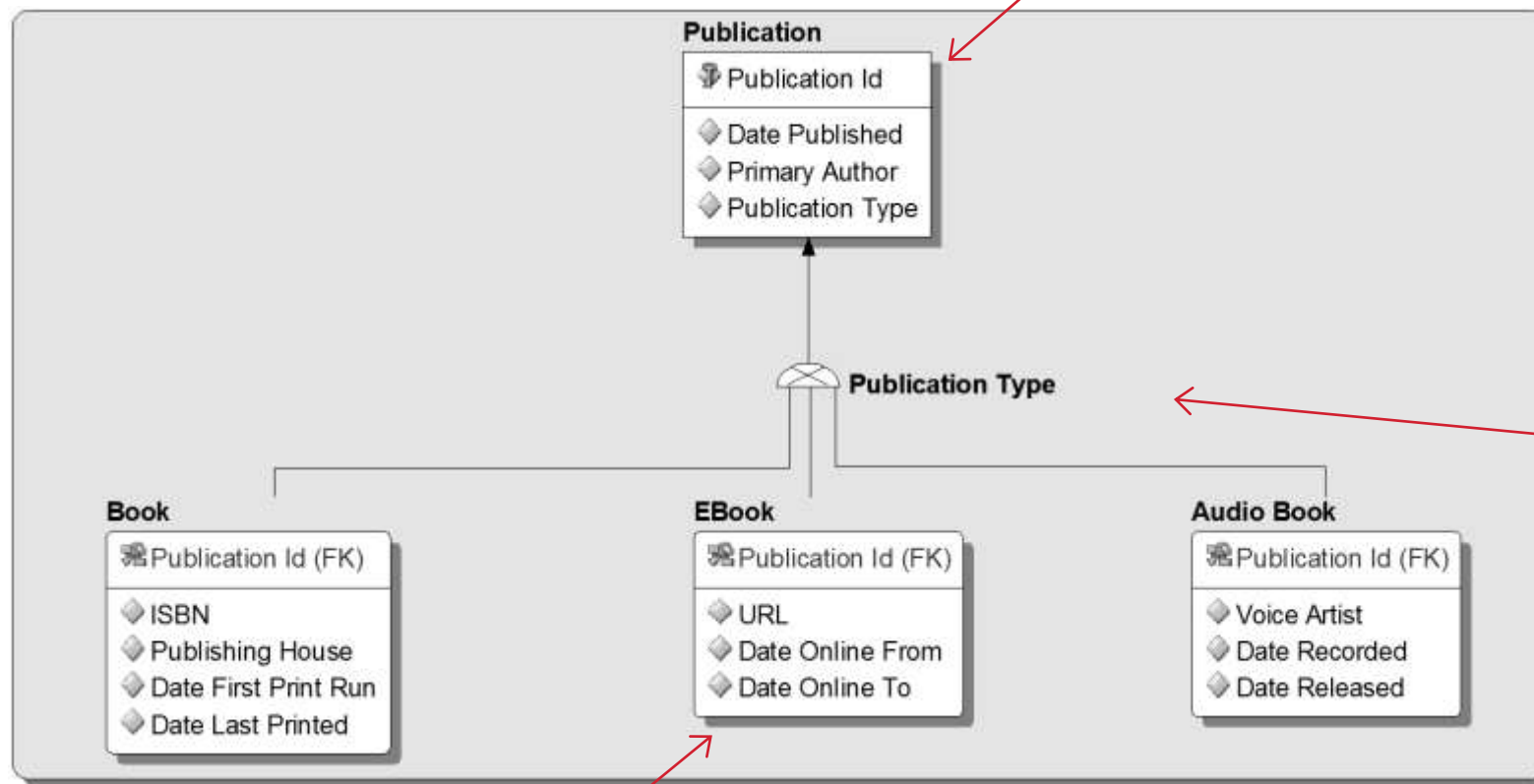
But we cannot have duplicated Attribute Names in an Entity hence the FK must have a role (e.g. Managed by).



It is possible to have **many** recursive relationships between the same entity and itself.



# Entity Subtypes



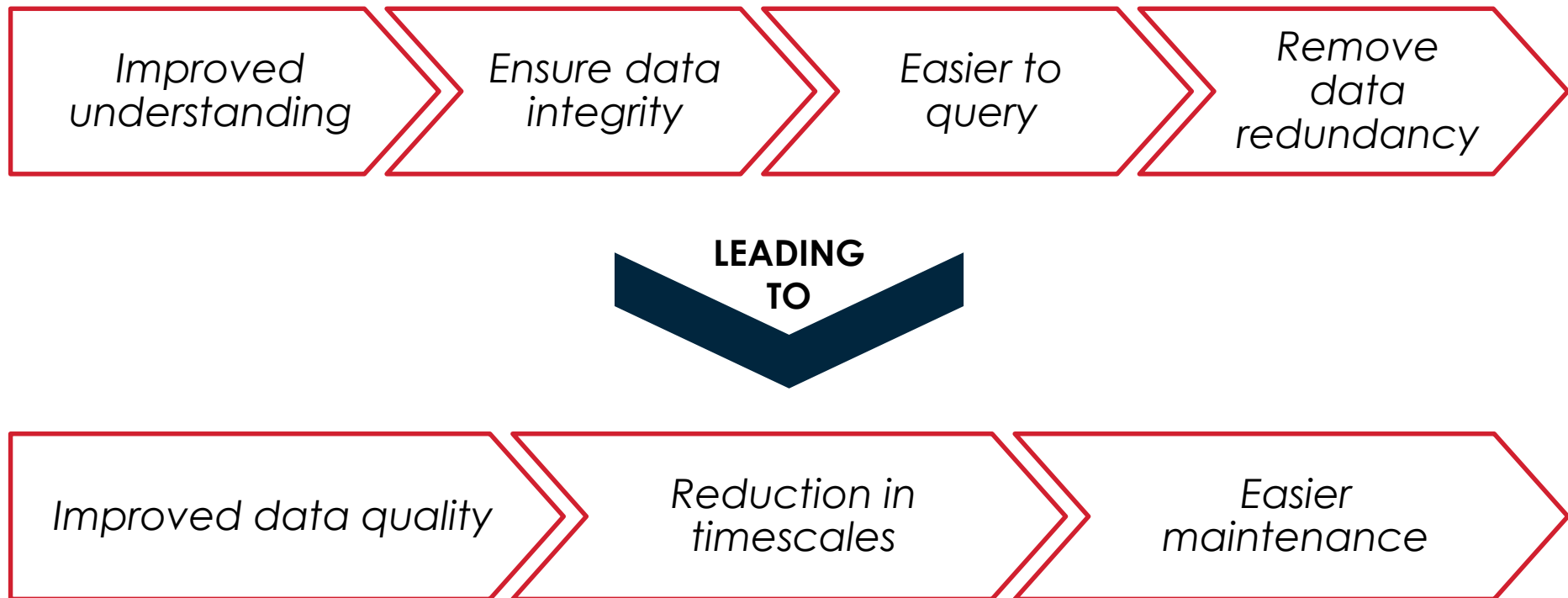
**Super-type Entity:**  
Contains the primary key and common attributes

**Discriminator:** Attribute to determine which subtype we are talking about

**Sub-type Entities:**  
Contain Specific attributes for each type

# Normalisation

## WHY NORMALISE A DATA MODEL?



# Normalisation Approaches

A data model is fully normalised when it is in Third Normal Form (3NF).

- 3NF is a normalisation method indicating 3 stages of normalisation:



Further normalisation methods can also be applied for very specific cases. See advanced course for details:

- Boyce/Codd normal form (BCNF)
- Fourth Normal Form (4NF)
- Fifth Normal Form (5NF)

# 1st Normal Form

## 1NF DEFINITION:

**EVERY NON-KEY ATTRIBUTE IN AN ENTITY MUST  
DEPEND ON IT'S PRIMARY KEY**

### A PRIMARY KEY MUST BE

- › **Unique** - the primary key uniquely identifies each instance of the entity
- › **Mandatory** – the primary key must be defined for every instance of the entity
- › **Unchanging** – while not mandatory, it is desirable that the primary key does not change

### TO PUT A MODEL INTO 1NF

- › Identify the primary key
- › Remodel repeating values
- › Remodel multi-valued attributes

# 1st Normal Form Example

NAME	GENDER	EMAIL ADDRESS
Barack Obama	Male	<a href="mailto:barack@whitehouse.org">barack@whitehouse.org</a> <a href="mailto:gobama@vote2012.com">gobama@vote2012.com</a> judgementday@dontnuke.com
David Cameron	Male	<a href="mailto:Callmedave@tory.co.uk">Callmedave@tory.co.uk</a> DC@uk.gov.com
Angela Merkel	Female	FrauAng@gov.de
Julia Gillard	Female	<a href="mailto:madeinwales@gov.au">madeinwales@gov.au</a> julia@outofwork.com

**To put this in  
first normal  
form we must:**



*Identify a  
primary key*



*Remodel the  
multi-valued  
attribute of Name*



*Remodel the multi-  
valued attribute of  
Email Address*

# 1st Normal Form Example

## A PRIMARY KEY MUST BE

- › **Unique** - the primary key uniquely identifies each instance of the entity
- › **Mandatory** – the primary key must be defined for every instance of the entity
- › **Unchanging** – while not mandatory, it is desirable that the primary key does not change

What is the primary key:  
Name,  
Gender, Email?

*So there is no primary key!  
So what do we do...?  
We create our own  
primary key (virtual key)*



**To put this in first normal form we must:**



*Identify a primary key*



*Remodel the multi-valued attribute of Name*



*Remodel the multi-valued attribute of Email Address*



# 1st Normal Form Example

NAME	GENDER	EMAIL ADDRESS
Barack Obama	Male	<a href="mailto:barack@whitehouse.org">barack@whitehouse.org</a> <a href="mailto:gobama@vote2012.com">gobama@vote2012.com</a> judgementday@dontnuke.com
David Cameron	Male	<a href="mailto:Callmedave@tory.co.uk">Callmedave@tory.co.uk</a> DC@uk.gov.com
Angela Merkel	Female	FrauAng@gov.de
Julia Gillard	Female	<a href="mailto:madeinwales@gov.au">madeinwales@gov.au</a> julia@outofwork.com

**Name = First Name + Last Name**

**To put this in first normal form we must:**



*Identify a primary key*



*Remodel the multi-valued attribute of Name*



*Remodel the multi-valued attribute of Email Address*

# 1st Normal Form Example

NAME	GENDER	EMAIL ADDRESS
Barack Obama	Male	<a href="mailto:barack@whitehouse.org">barack@whitehouse.org</a> <a href="mailto:gobama@vote2012.com">gobama@vote2012.com</a> judgementday@dontnuke.com
David Cameron	Male	<a href="mailto:Callmedave@tory.co.uk">Callmedave@tory.co.uk</a> DC@uk.gov.com
Angela Merkel	Female	FrauAng@gov.de
Julia Gillard	Female	<a href="mailto:madeinwales@gov.au">madeinwales@gov.au</a> julia@outofwork.com

Next we look at this

**To put this in first normal form we must:**



*Identify a primary key*



*Remodel the multi-valued attribute of Name*



*Remodel the multi-valued attribute of Email Address*

# 1NF: Email Address?

- › A multi-valued attribute: name & domain?
- › Can we identify ALL available types?
  - » Home Email
  - » Work Email
  - » Club Email
- › An exhaustive list?
- › No; a Person can have any number of email addresses.
- › We need to allow for a Person having any number of email addresses.

*Is it just a  
compound  
attribute?*

*No, it is  
multi-  
valued!*

**To put this in  
first normal  
form we must:**



*Identify a  
primary key*



*Remodel the  
multi-valued  
attribute of Name*



*Remodel the multi-  
valued attribute of  
Email Address*

# 1st Normal Form Example

## 1NF DEFINITION:

**EVERY NON-KEY ATTRIBUTE IN AN ENTITY MUST DEPEND ON IT'S PRIMARY KEY**

### PERSON

PERSON ID	FIRST NAME	LAST NAME	GENDER
1	Barack	Obama	Male
2	David	Cameron	Male
3	Angela	Merkel	Female
4	Julia	Gillard	Female

***& as it's a simple example  
they are in 2NF & 3NF too!***

### PERSON EMAIL

PERSON ID	EMAIL ADDRESS
1	<a href="mailto:barack@whitehouse.org">barack@whitehouse.org</a>
1	<a href="mailto:gobama@vote2012.com">gobama@vote2012.com</a>
1	<a href="mailto:judgementday@dontnuke.com">judgementday@dontnuke.com</a>
2	<a href="mailto:Callmedave@tory.co.uk">Callmedave@tory.co.uk</a>
2	<a href="mailto:DC@uk.gov.com">DC@uk.gov.com</a>
3	<a href="mailto:FrauAng@gov.de">FrauAng@gov.de</a>
4	<a href="mailto:madeinwales@gov.au">madeinwales@gov.au</a>
4	<a href="mailto:julia@outofwork.com">julia@outofwork.com</a>

We have now put our politicians in 1NF!

# 1st Normal Form Example

## 1NF DEFINITION:

**EVERY NON-KEY ATTRIBUTE IN AN ENTITY MUST  
DEPEND ON IT'S PRIMARY KEY**

REGISTRATION	MODEL	CHASSIS NUMBER	MILEAGE	FEATURES
HV62SYG	Lexus 450H	76365296745568432	7,129	Electric Windows Satellite Navigation Bluetooth integration Head Up Display Speech Control
Y612 SYG	Audi A4	13847621837653275	10,732	Electric Windows Bluetooth integration
WN09 UTS	BMW 320d	32178468273647327	31,123	Electric Windows Satellite Navigation Bluetooth integration
WU52XUX	Ford Focus	71283459735474924	104,123	Electric Windows

**Turn this  
into 1NF:  
State your  
Keys**

# 1st Normal Form Example

## CAR ENTITY

CHASSIS NUMBER	REGISTRATION	MANUFACTURER	MODEL	MILEAGE
76365296745568432	HV62SYG	LEXUS	RX450H	7,129
13847621837653275	Y612 SYG	AUDI	A4	10,732
32178468273647327	WN09 UTS	BMW	320D	31,123
71283459735474924	WU52XUX	FORD	FOCUS	104,123

## CAR FEATURE ENTITY

CHASSIS NUMBER	FEATURE
76365296745568432	Electric Windows
76365296745568432	Satellite Navigation
76365296745568432	Bluetooth integration
76365296745568432	Head Up Display
76365296745568432	Speech Control
13847621837653275	Electric Windows
13847621837653275	Bluetooth integration
32178468273647327	Electric Windows
32178468273647327	Satellite Navigation
32178468273647327	Bluetooth integration
71283459735474924	Electric Windows

The primary keys are in **BLUE**



But, refer back to primary key criteria



Feature names are likely to change

# 1st Normal Form Example

## CAR ENTITY

CHASSIS NUMBER	REGISTRATION	MANUFACTURER	MODEL	MILEAGE
76365296745568432	HV62SYG	LEXUS	RX450H	7,129
13847621837653275	Y612 SYG	AUDI	A4	10,732
32178468273647327	WN09 UTS	BMW	320D	31,123
71283459735474924	WU52XUX	FORD	FOCUS	104,123

## CAR FEATURE ENTITY

CHASSIS NUMBER	FEATURE ID	FEATURE
76365296745568432	1	Electric Windows
76365296745568432	2	Satellite Navigation
76365296745568432	3	Bluetooth integration
76365296745568432	4	Head Up Display
76365296745568432	5	Speech Control
13847621837653275	1	Electric Windows
13847621837653275	3	Bluetooth integration
32178468273647327	1	Electric Windows
32178468273647327	2	Satellite Navigation
32178468273647327	3	Bluetooth integration
71283459735474924	1	Electric Windows



# 2nd Normal Form

## 2NF DEFINITION:

**EACH ENTITY MUST HAVE THE FEWEST POSSIBLE  
CORRECT PRIMARY KEY ATTRIBUTES**

*For each non-  
key attribute  
(i.e. not a primary,  
foreign or alternate  
key)*

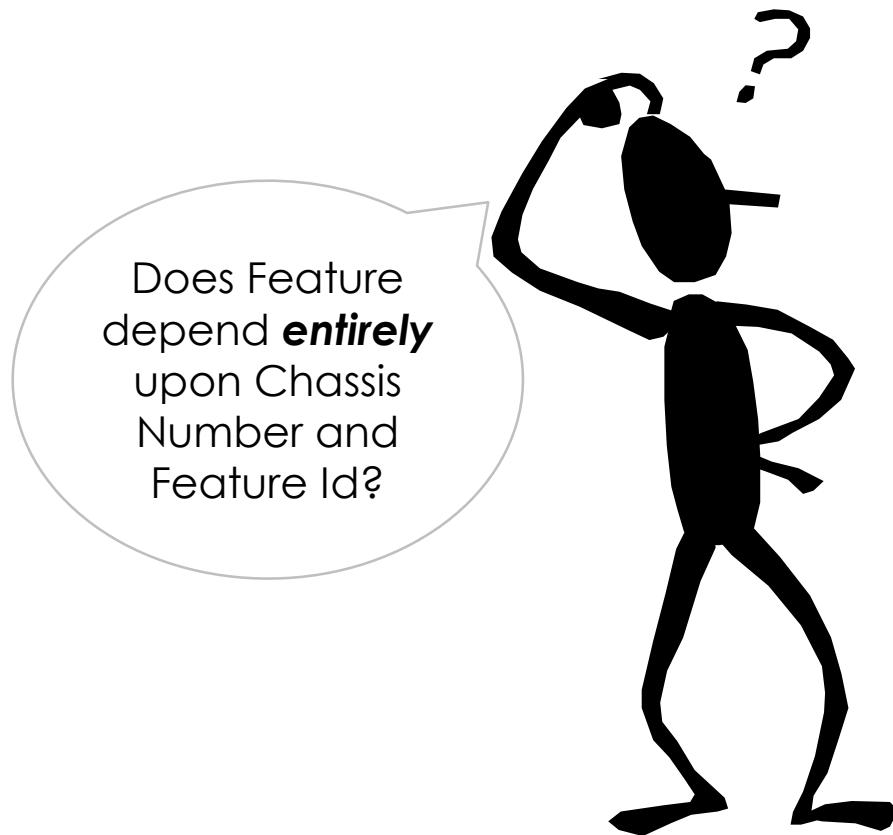


*Test if it  
depends  
**entirely** on the  
primary key*



*If it doesn't,  
move it out to a  
new entity*

# 2nd Normal Form Example




## CAR FEATURE ENTITY

CHASSIS NUMBER	FEATURE ID	FEATURE
76365296745568432	1	Electric Windows
76365296745568432	2	Satellite Navigation
76365296745568432	3	Bluetooth integration
76365296745568432	4	Head Up Display
76365296745568432	5	Speech Control
13847621837653275	1	Electric Windows
13847621837653275	3	Bluetooth integration
32178468273647327	1	Electric Windows
32178468273647327	2	Satellite Navigation
32178468273647327	3	Bluetooth integration
71283459735474924	1	Electric Windows

# 2nd Normal Form Example

*Car Feature is an associative entity*



**CAR ENTITY**

CHASSIS NUMBER	REGISTRATION	MANUFACTURER	MODEL	MILEAGE
76365296745568432	HV62SYG	LEXUS	RX450H	7,129
13847621837653275	Y612 SYG	AUDI	A4	10,732
32178468273647327	WN09 UTS	BMW	320D	31,123
71283459735474924	WU52XUX	FORD	FOCUS	104,123

**FEATURE ENTITY**

FEATURE ID	FEATURE
1	Electric Windows
2	Satellite Navigation
3	Bluetooth integration
4	Head Up Display
5	Speech Control

**CAR FEATURE ENTITY**

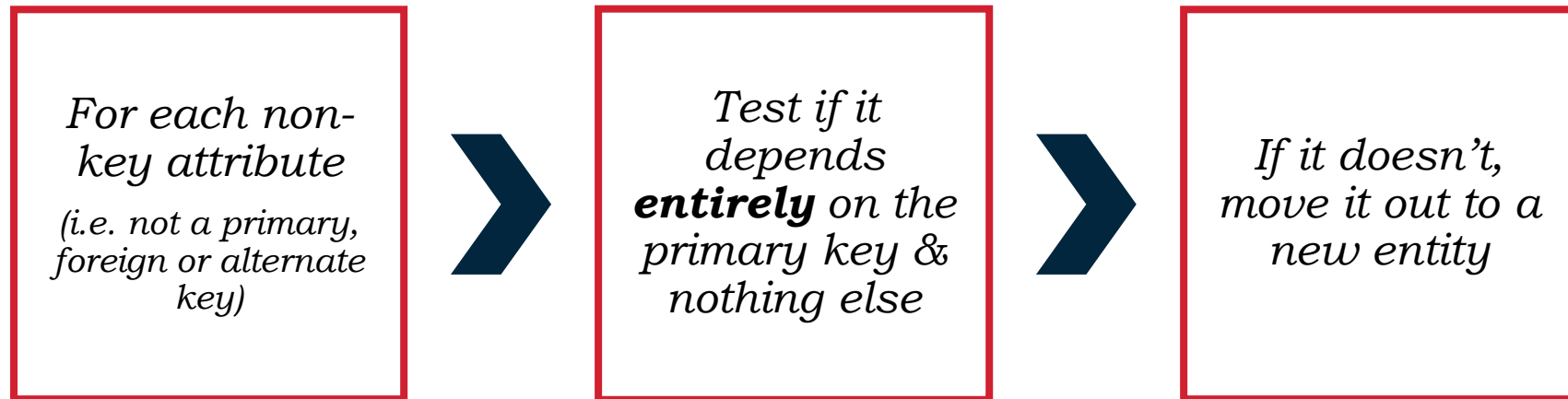
CHASSIS NUMBER	FEATURE ID
76365296745568432	1
76365296745568432	2
76365296745568432	3
76365296745568432	4
76365296745568432	5
13847621837653275	1
13847621837653275	3
32178468273647327	1
32178468273647327	2
32178468273647327	3
71283459735474924	1

We have now put our cars into 2NF

# 3rd Normal Form

## 3NF DEFINITION:

EACH NON KEY ELEMENT MUST BE DIRECTLY  
DEPENDENT UPON THE PRIMARY KEY AND NOT  
UPON ANY OTHER NON-KEY ATTRIBUTES



# 3rd Normal Form Example

CHASSIS NUMBER	REGISTRATION	MANUFACTURER	MODEL	MILEAGE
76365296745568432	HV62SYG	LEXUS	RX450H	7,129
13847621837653275	Y612 SYG	AUDI	A4	10,732
32178468273647327	WN09 UTS	BMW	320D	31,123
71283459735474924	WU52XUX	FORD	FOCUS	104,123

*For 3NF, all attributes must depend only on Chassis Number.*



*But “Model” also depends upon Manufacturer*

# 3rd Normal Form Example

CHASSIS NUMBER	REGISTRATION	MODEL ID	MILEAGE
76365296745568432	HV62SYG	1	7,129
13847621837653275	Y612 SYG	2	10,732
32178468273647327	WN09 UTS	3	31,123
71283459735474924	WU52XUX	4	104,123

MODEL ID	MANUFACTURER ID	MODEL
1	1	RX450H
2	2	A4
3	3	320D
4	4	FOCUS

***The model name on its own is not a key candidate since it may not be unique:***

*Its possible that 2 manufacturers may make a car with the same name. The manufacturer and model together make a key.*

# 3rd Normal Form Example

## 3NF DEFINITION:

**EACH NON KEY ELEMENT MUST BE DIRECTLY DEPENDENT UPON THE PRIMARY KEY AND NOT UPON ANY OTHER NON-KEY ATTRIBUTES**

CHASSIS NUMBER	REGISTRATION	MODEL ID	MILEAGE
76365296745568432	HV62SYG	1	7,129
13847621837653275	Y612 SYG	2	10,732
32178468273647327	WN09 UTS	3	31,123
71283459735474924	WU52XUX	4	104,123

MODEL ID	MANUFACT-URER ID	MODEL
1	1	RX450H
2	2	A4
3	3	320D
4	4	FOCUS

MANUFACT-URER ID	MANUFACT-URER NAME	CONTACT EMAIL
1	Lexus	hitori@lexus.jp
2	Audi	hans@audi.de
3	BMW	woflgang@bmw.de
4	Ford	dwane@ford.com

We have now put our cars into 3NF



# Normalisation Summary

0NF 


---

1NF 

---

2NF 

---

3NF 

---

4& 5NF 

---

Un-normalised (UNF or 0NF)  
Contains a “repeating group”

First Normal Form (1NF)  
Repeating attributes moved down to associative entities

Second Normal Form (2NF)  
Only applies to dependent entities  
No attributes in a child entity are really facts about a parent (or grandparent). No characteristic or associative entity redundantly contains facts from its parent(s) – if it does, move the fact(s) up and if necessary create a new parent entity

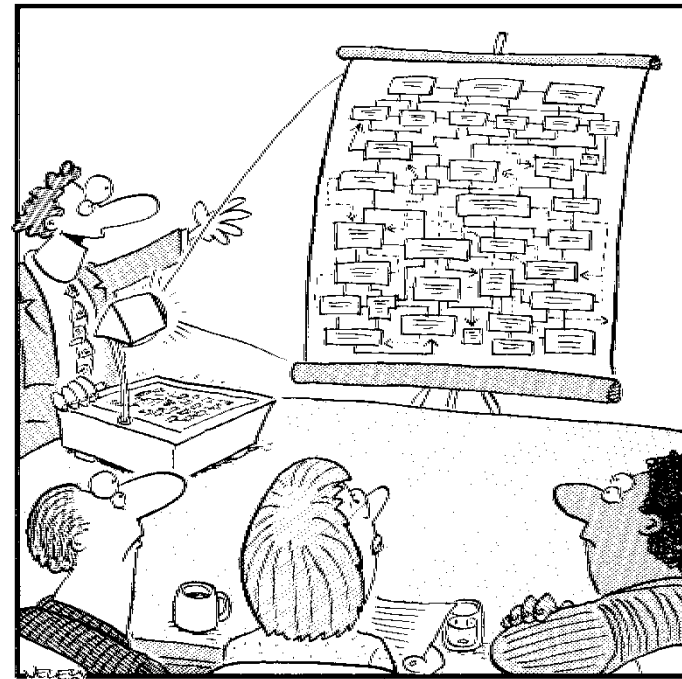
Third Normal Form (3NF)  
If any entity redundantly contains facts from a related (non-parent) entity, move the fact(s) out to the other entity and create a new entity if necessary

Fourth and Fifth Normal Form (4NF, 5NF)  
“Large” (3-way or more) associatives need to be broken down into more granular entities

# Graphical Principles

## OUR MODELS SHOULD AID UNDERSTANDING BY:

- › Using visual cues consistently
- › Having a starting point and direction
- › Abstracting
- › Masking unnecessary detail
- › Highlighting what matters



*"Let's start here with Special Tax Rate Variation Comment Type..."*

# Dimensional Data Models

Designed for the rapid recovery of information to be delivered to OLAP systems or ad hoc analysis



Modeled based on the dimensional modeling principles popularised by Ralph Kimball

*“Entity-relationship modelling is a logical design technique that seeks to eliminate data redundancy”*

**A good choice for  
OLTP systems**

*“Dimensional modelling is a design technique that seeks to present data in a way that maximises both ease of use and query performance.”*

**A good choice for Data  
Warehouses / Business  
Intelligence systems**

# Model Features

## RELATIONAL

- › Optimised for OLTP
- › Normalised
- › Low redundancy
- › Relationships between entities are explicit
- › Tightly coupled to business model

## DIMENSIONAL

- › Optimised for reporting
- › Business Entities are denormalised
- › More redundancy to support faster query performance
- › Relationships between entities are implicit
- › Loosely coupled to business model

# A Dimensional Model

## 'STAR SCHEMA'

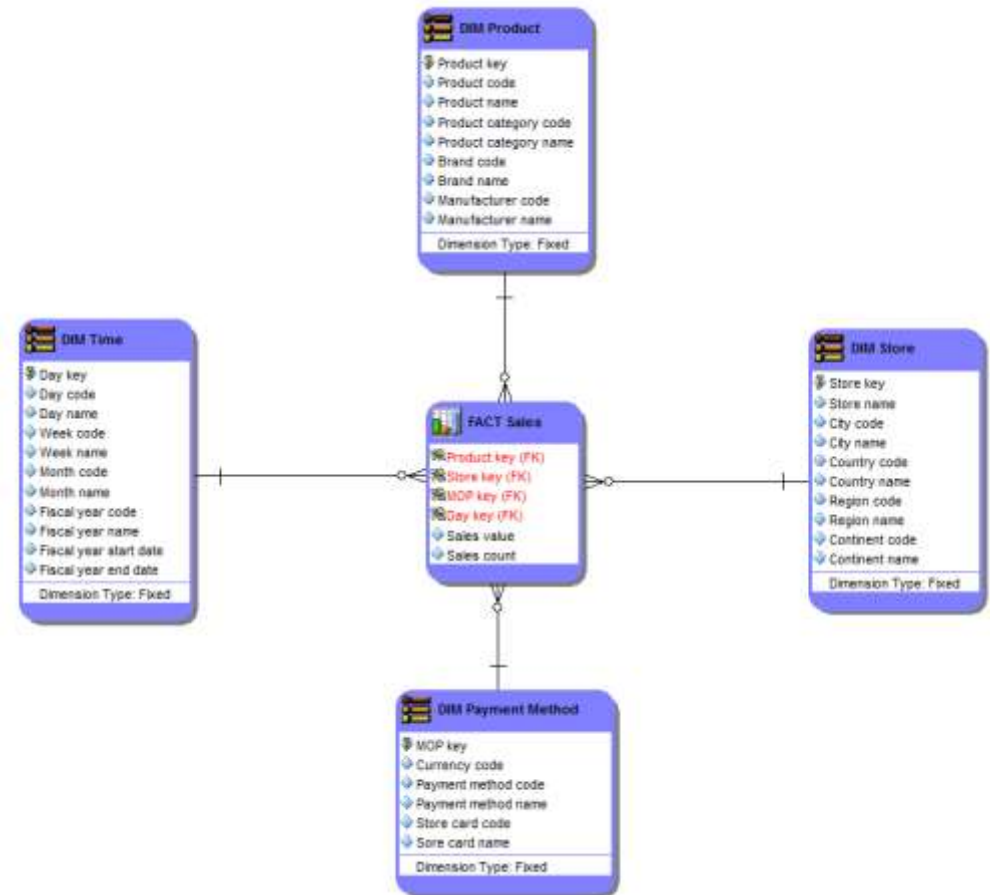
composed of Dimension and Fact tables

### Dimension tables

- › EG: Location, Product, Time, Promotion, Organisation ...
- › Product dimension includes Product Type, Brand, Manufacturer
- › Store dimension includes Country, Continent

### Fact tables

- › Contains measures (e.g. Sales Value) and dimension FK's
- › Dimension columns are FK's pointing to the respective dimensions.



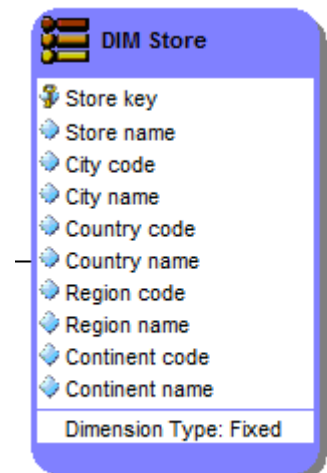
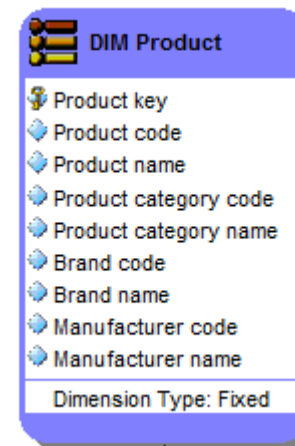
# Dimensions & Hierarchies

Hierarchies for the dimensions are stored in the dimensional table itself so there is no need for the individual hierarchical lookup tables be shown in the model.

Records in dimension tables correspond to nouns, the tables are “short” – 10s to 1,000s of records

Rich set of attributes, tables are “wide” – many columns & the data changes slowly

Denormalised so no need to join to further lookup tables. This means there is some redundancy



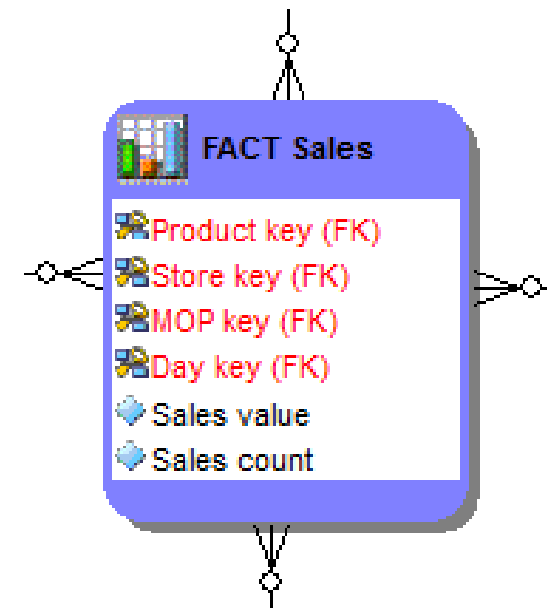
# Fact Tables

Records in fact tables correspond to **events, transactions, or measurements.**

Data is added regularly; the tables are “long” – often millions of records

Rich set of attributes; the tables are “narrow” – minimal number of columns

Low redundancy



The most useful measures are “additive”



# Advanced Concepts

Aggregates



Bridge  
Tables

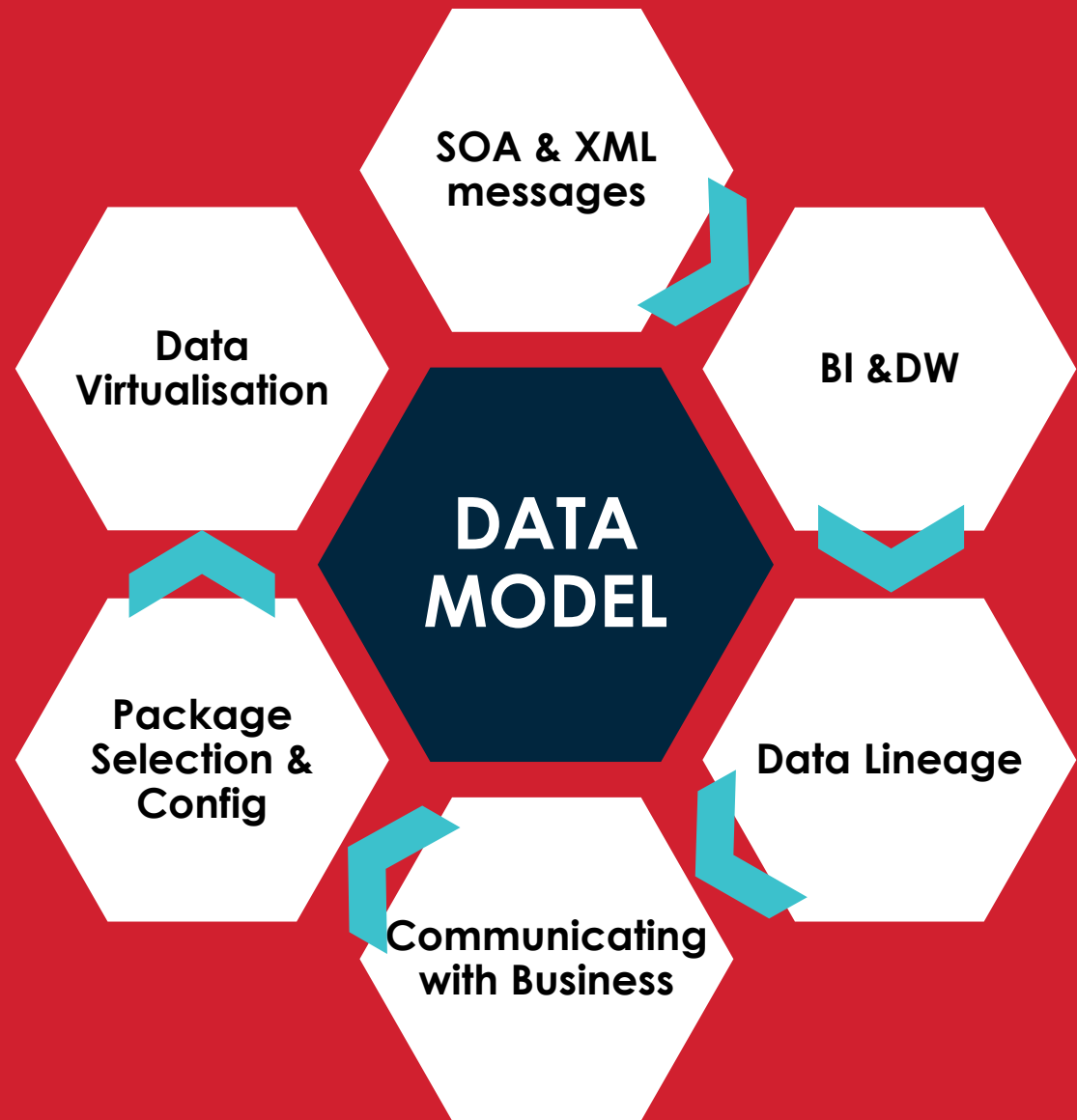


Slowly  
Changing  
Dimensions

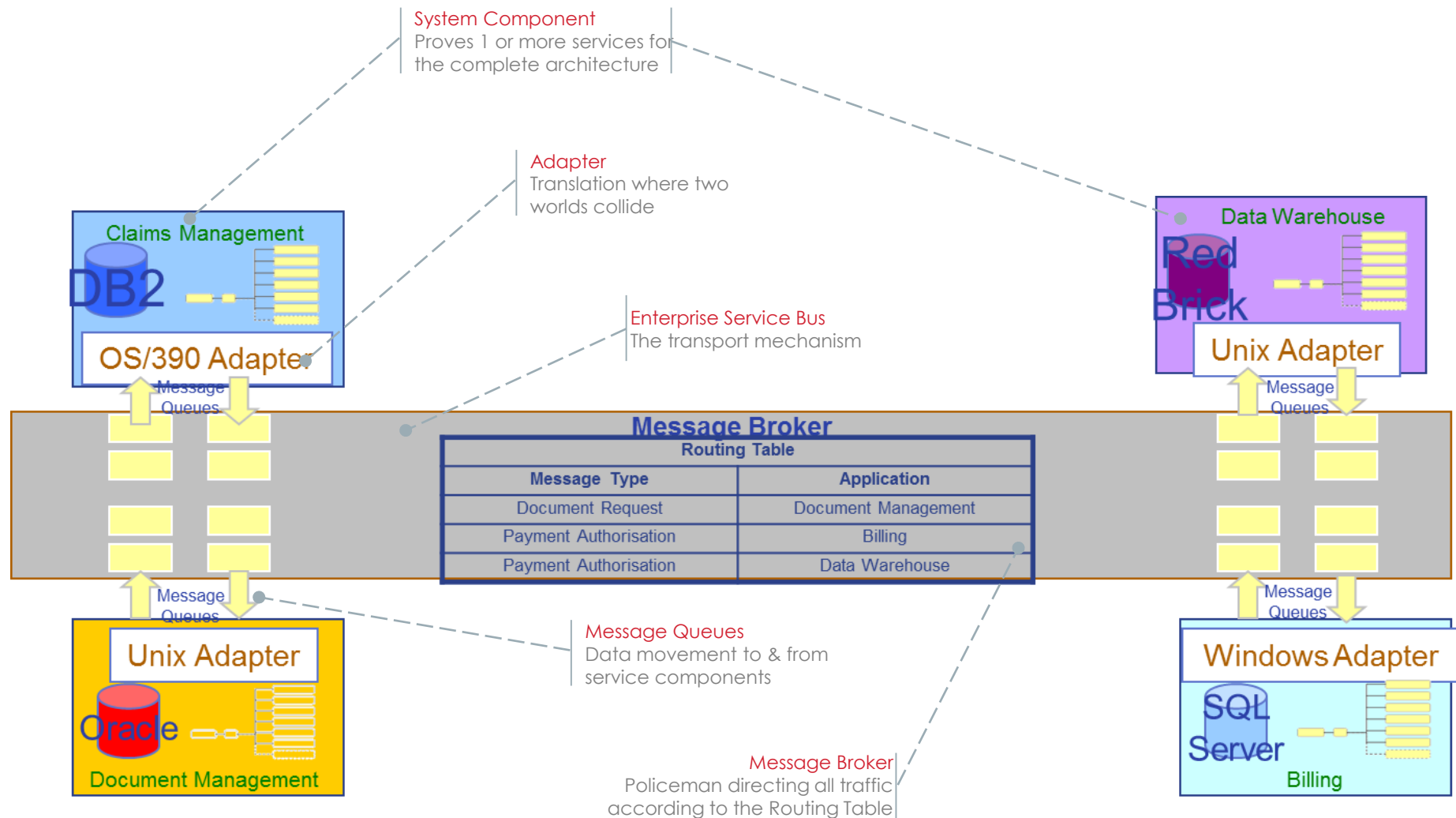


Factless  
Facts

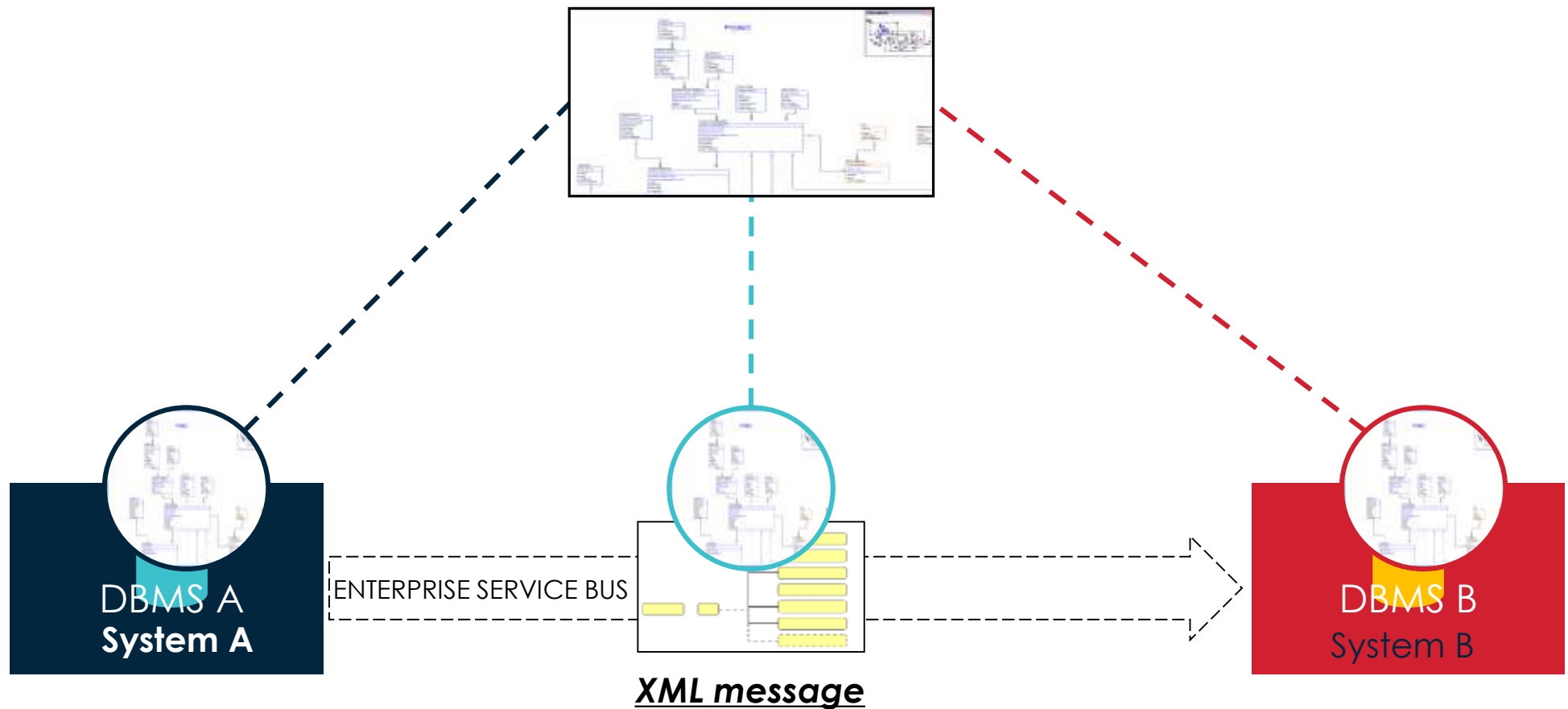
# Data Modelling: It's NOT just for DataBase Design



# SOA 101



# XML Messages Need Data Models



# XML versus E/R Structures

## XML

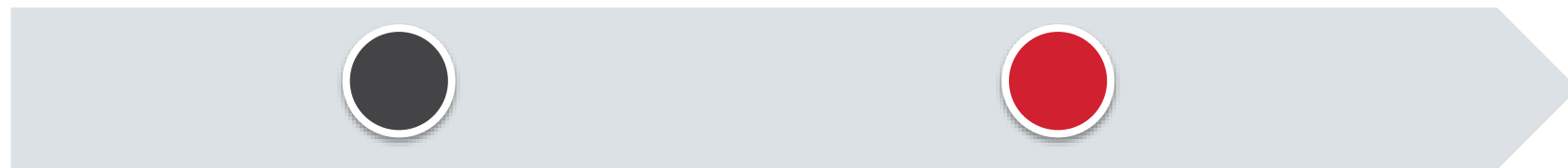
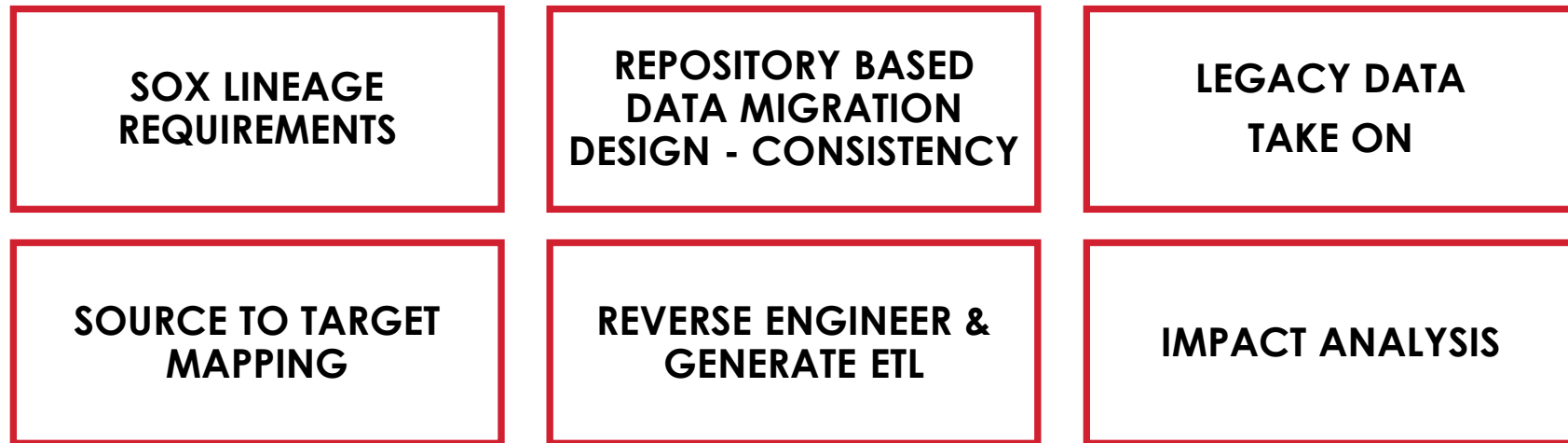
- › Hierarchical - tree structure.
- › Each entity has just one parent.
- › Used for transfer of data.
- › Shared data appears multiple times in multiple messages.

## E/R Structure

- › Relational - network structure.
- › Each entity can have many parents.
- › Used for storage and maintenance of data.
- › Shared data typically appears just once.

Well defined processes for converting  
ER/Model into XML  
Anyone Remember IMS & DL/1?

# Data Lineage



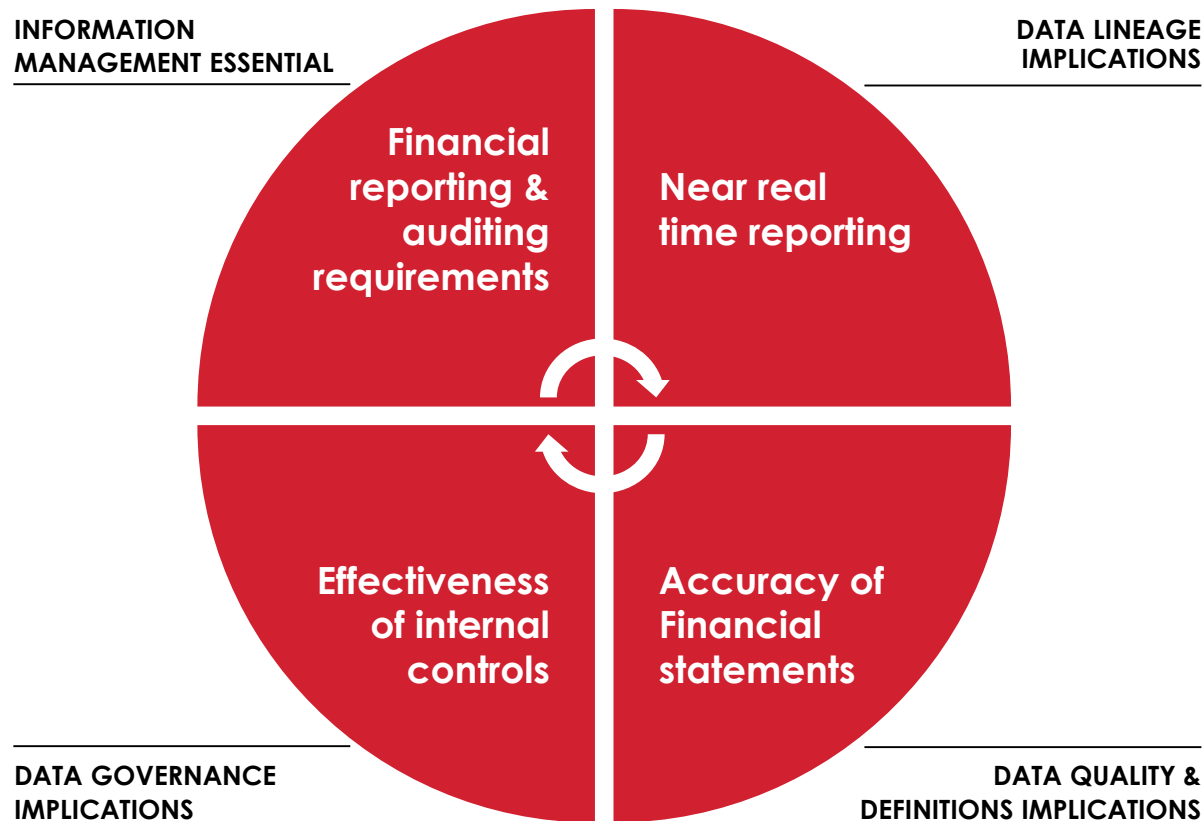
## TRANSFORMATIONS

- *What has been done to the data?*

## BUSINESS PROCESSES

- *Which business processes can be applied to the data?*
- *What type of actions do those processes perform (Create, Read, Update, Delete)?*
- *Audit Trail – who has supplied, accessed, updated, approved and deleted the data and when?*  
*Which processes have acted on the data?*

# Data Lineage: e.g. SOX



Sarbox has 66 sections of which three will probably have the greatest impact on content, document and process technologies.

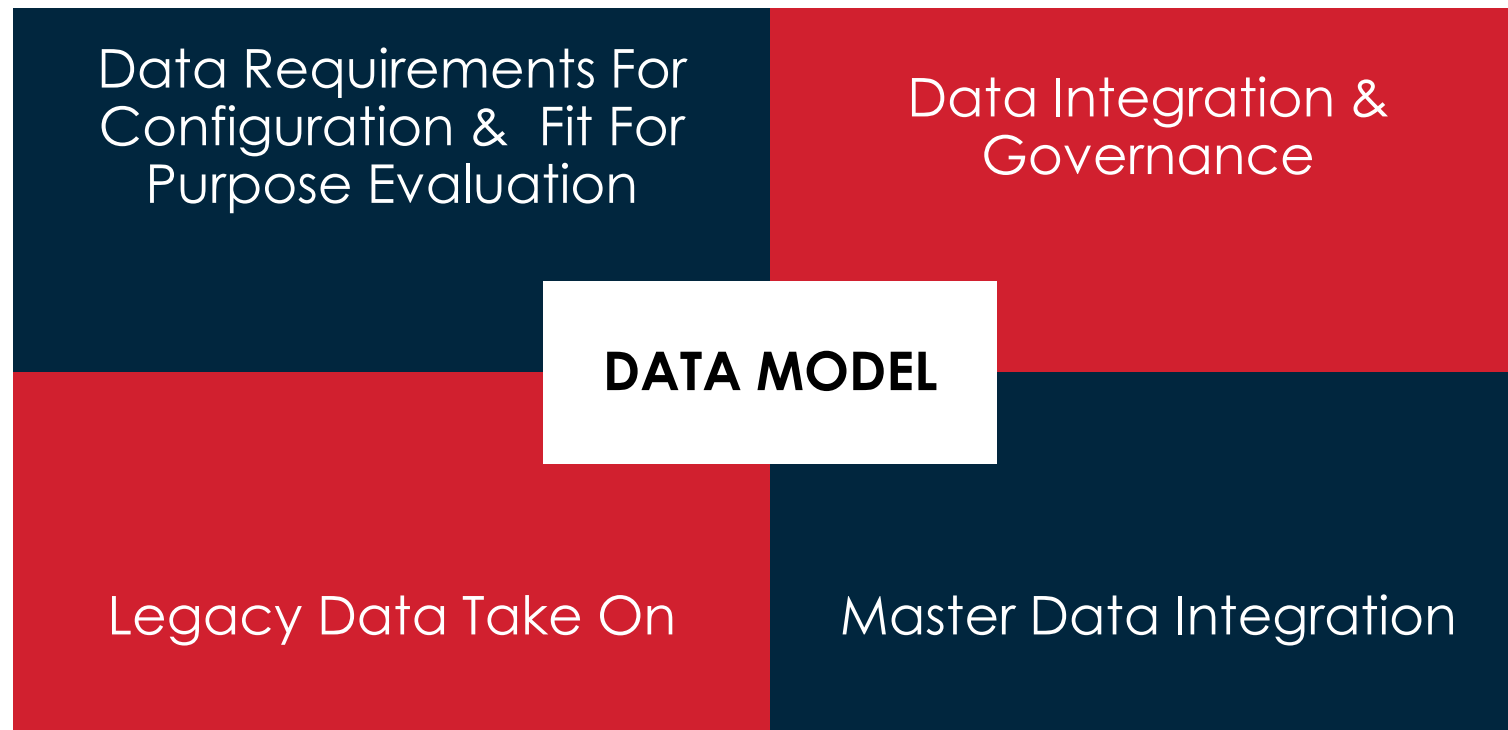
- **Section 302:** Requires CEOs and CFOs of public companies to attest quarterly and annually on their responsibilities over the material accuracy of the company's financial statements and relating internal controls.

- **Section 404:** Requires public companies to report annually on the effectiveness of their internal controls and requires their external auditors to attest on management's conclusions.

- **Section 409:** Requires near real-time reporting for public companies. The SEC has proposed a 48-hour deadline for filing an 8-K, a notice of material change.



# Package / ERP systems



# Data Modelling For Packages / ERP Systems

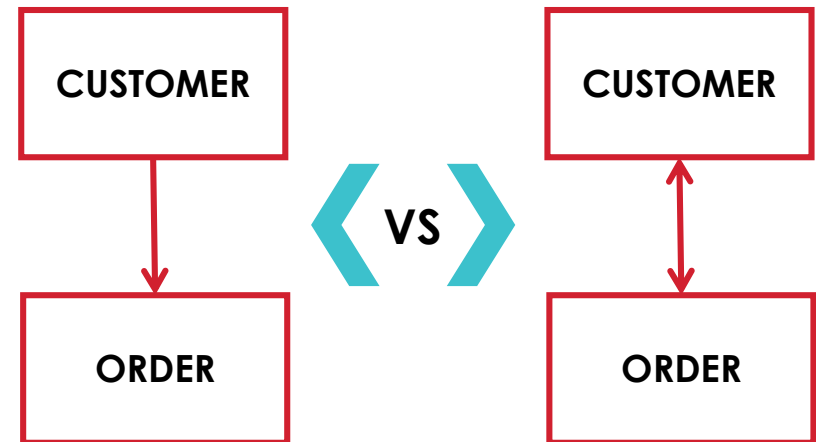
For requirements gathering ... *But what if we've got to use package X?*

Identifying gaps

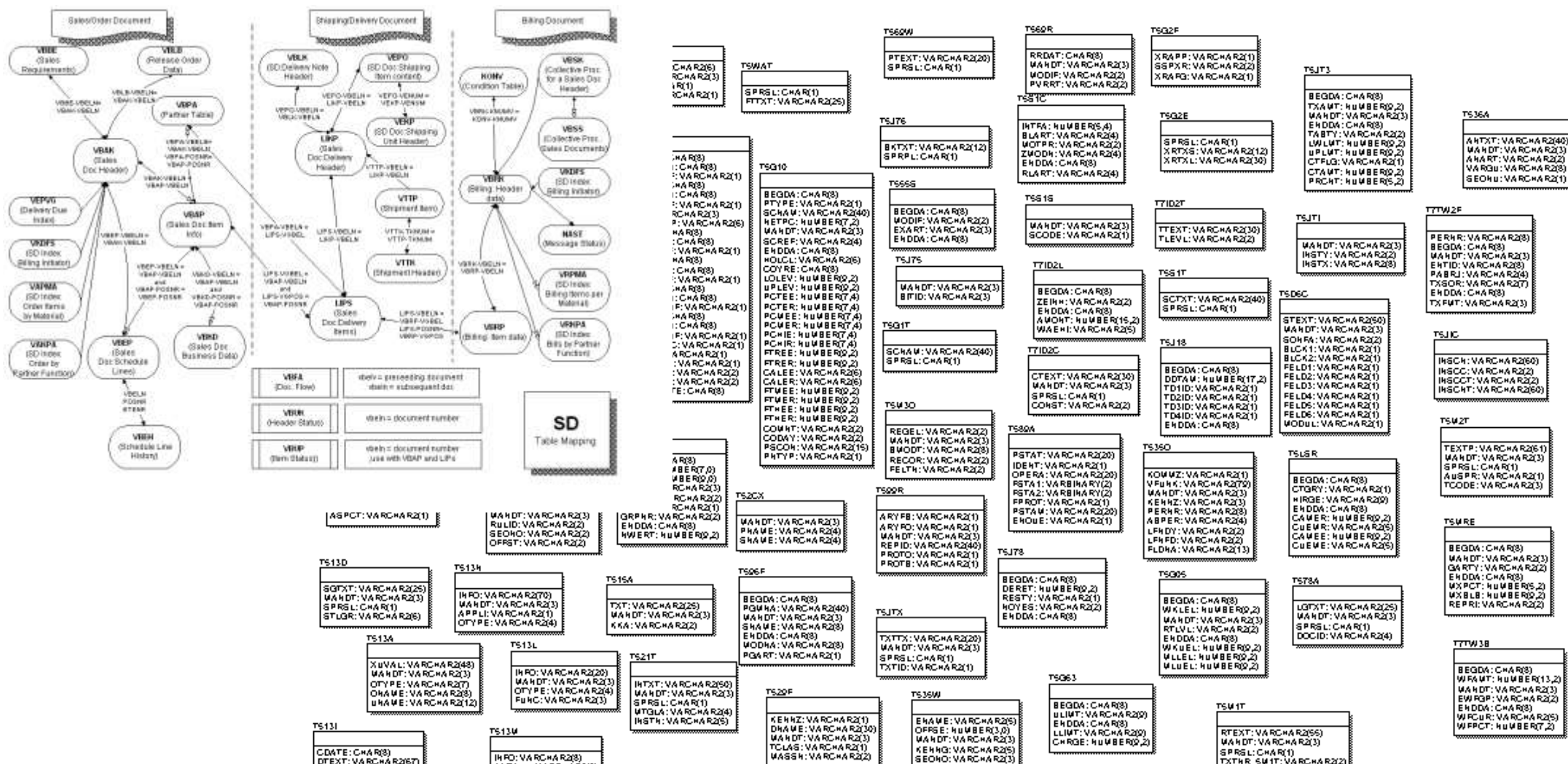
For Data migration / take on

Master Data alignment

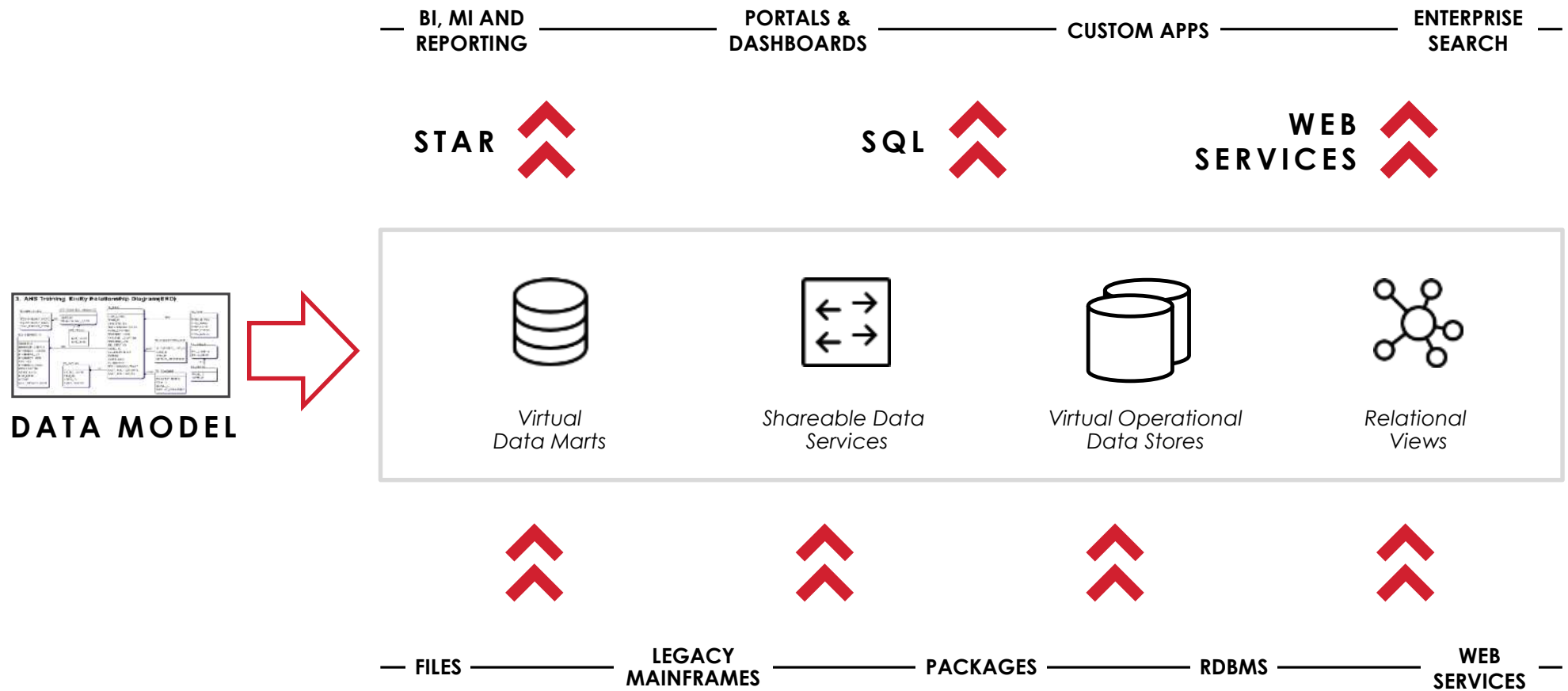
Data lineage (particularly important with Data Lineage & SOX compliance issues)



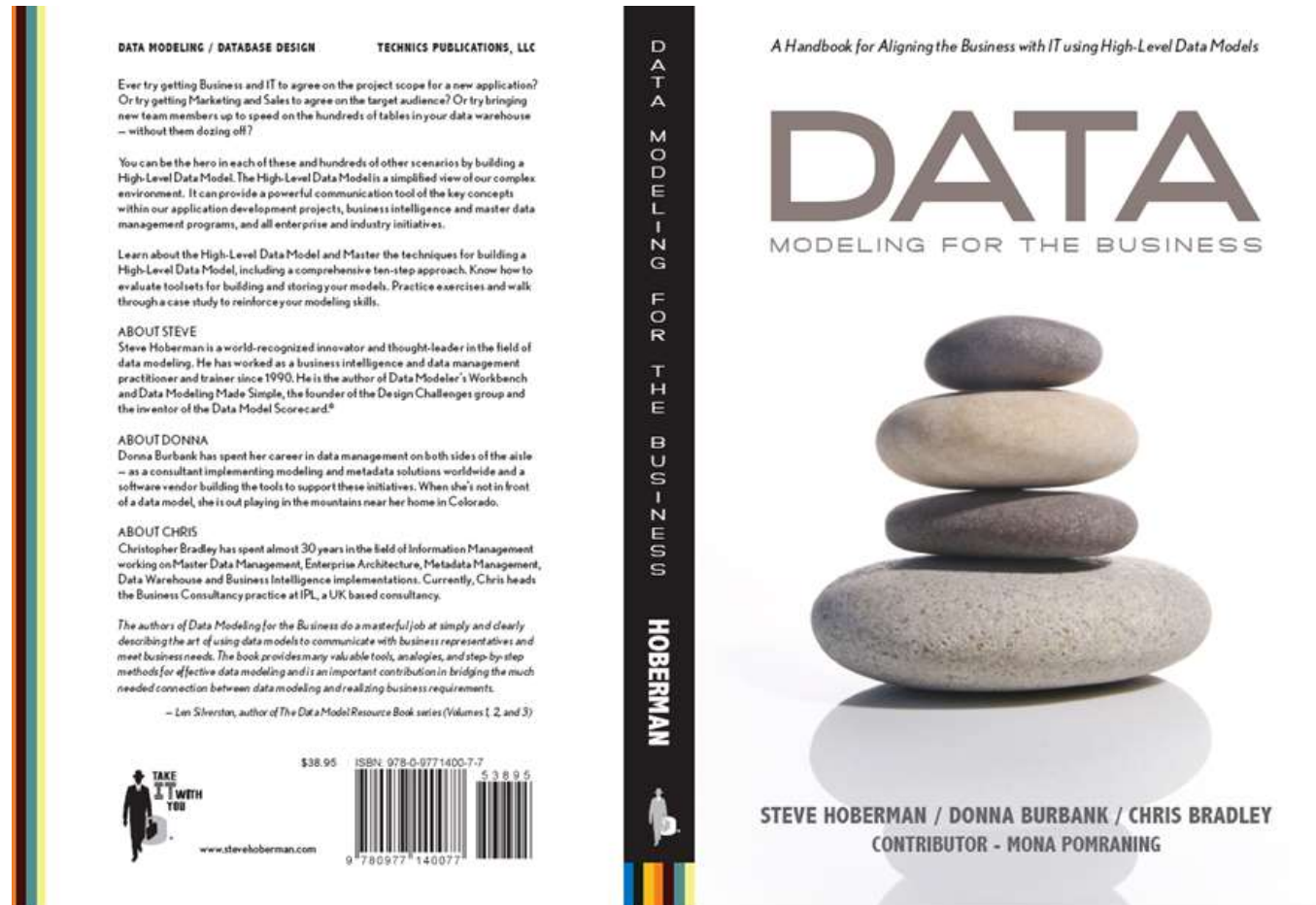
# You Can't Always Reverse Engineer



# Data Virtualisation



# Communicating With The Business



# **Data Modelling does not have to be Complicated!**

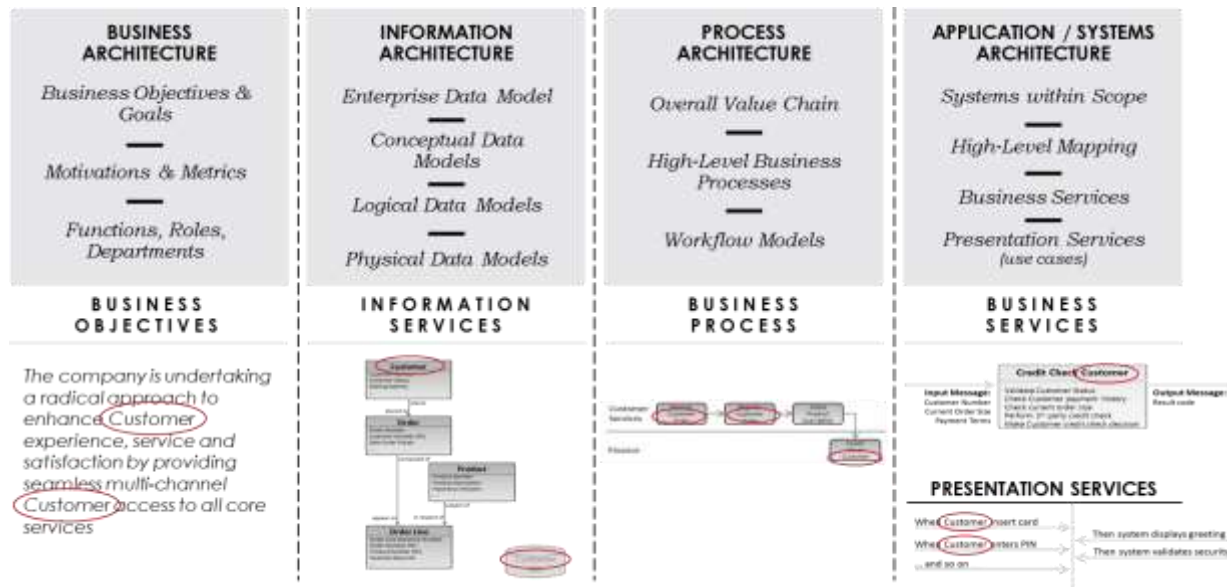
**If you can write a sentence, you can build a data model.**

**If you understand how your business works, you can build a data model.**

**Businesspeople should be involved in the development of data models, because only they understand the business needs and rules.**

**Understanding data modelling basics will help the Business better communicate with IT**

# Summary



*All of the Architecture disciplines use the language (and rules) of the data model*

Data is at the heart of ALL architecture disciplines

Data has to be understood to be managed

Different levels of models for different purposes

It's NOT just for DBMS design

Data models are not (just) art

Professional development: certification & training



# Christopher Bradley

INFORMATION MANAGEMENT STRATEGIST



Chris.Bradley@DMAAdvisors.co.uk



+44 7973 184475



@inforacer



uk.linkedin.com/in/christophermichaelbradley/



infomanagementlifeandpetrol.blogspot.com



The background of the slide features three iconic Eames-style chairs in red, blue, and yellow, arranged in a row on a light-colored wooden floor against a plain white wall. The chairs have a distinctive molded plastic shell and a chrome-plated metal wire base.

*Data Modelling 101*

**THANK YOU**

**Christopher Bradley**  
**Information Strategist**