Avancier Methods (AM)
Data Architecture

Enterprise level data architecture
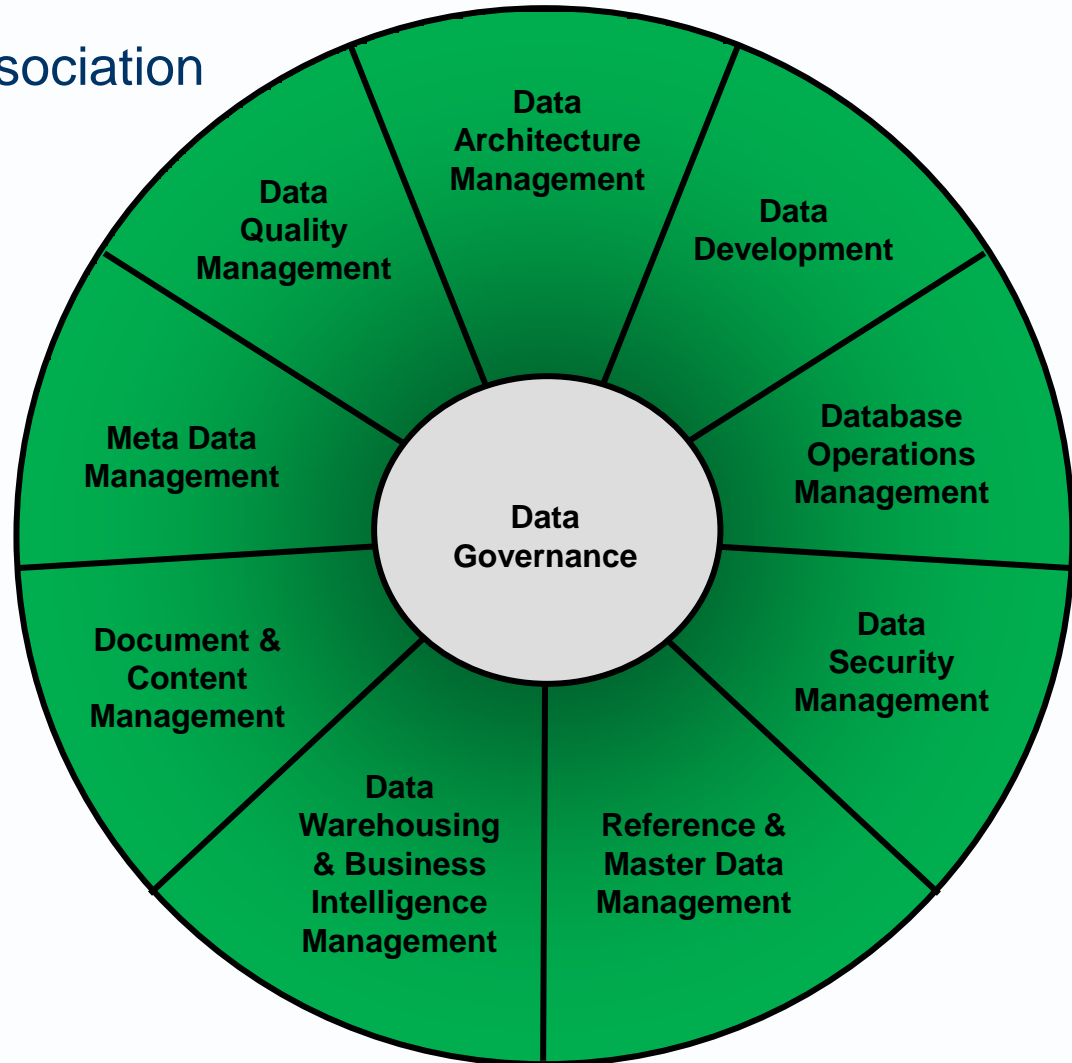
Enterprise Information Architecture

► Leading bank is urgently seeking a proven EA to engage and lead IT projects including the **Enterprise Information Architecture**

■ information models and flows,

■ data dictionaries, data standards

■ data quality standards and processes

► develop and maintain the logical **Enterprise Information Architecture** that enables seamless information interoperability of all Bank systems for efficiency and cost-effectiveness.

► eutopiaonline.com

Enterprise Information Architecture?

- ► The Data Management Association
- ► (DAMA)
- ► define its scope thus:

Avancier

DAMA principles

DAMA's

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable
- **Data is an asset with unique properties**
- **The value of data can and should be expressed in economic terms**

► Consider the value of data to a business
► Address data qualities
► Define data using data
► Take a business-wide view
► Consider data over its life time

Data Management Requirements are Business Requirements
- **Managing data means managing the quality of data**
- **It takes Metadata to manage data**
- **It takes planning to manage data**
- **Data management requirements must drive Information Technology decisions**

Data Management depends on diverse skills
- **Data management is cross-functional**
- **Data management requires an enterprise perspective**
- **Data management must account for a range of perspectives**

Data Management is lifecycle management
- **Different types of data have different lifecycle characteristics**
- **Managing data includes managing the risks associated with data**

In short, data architects are concerned with

► Data at rest   in store, in memory
► Data in motion   in flows, in messages
► Data qualities   confidentiality, integrity and availability (CIA)

Manufacturing   **Quality**   **Commercials**

Supply Chain   **Safety**   **Regulatory**
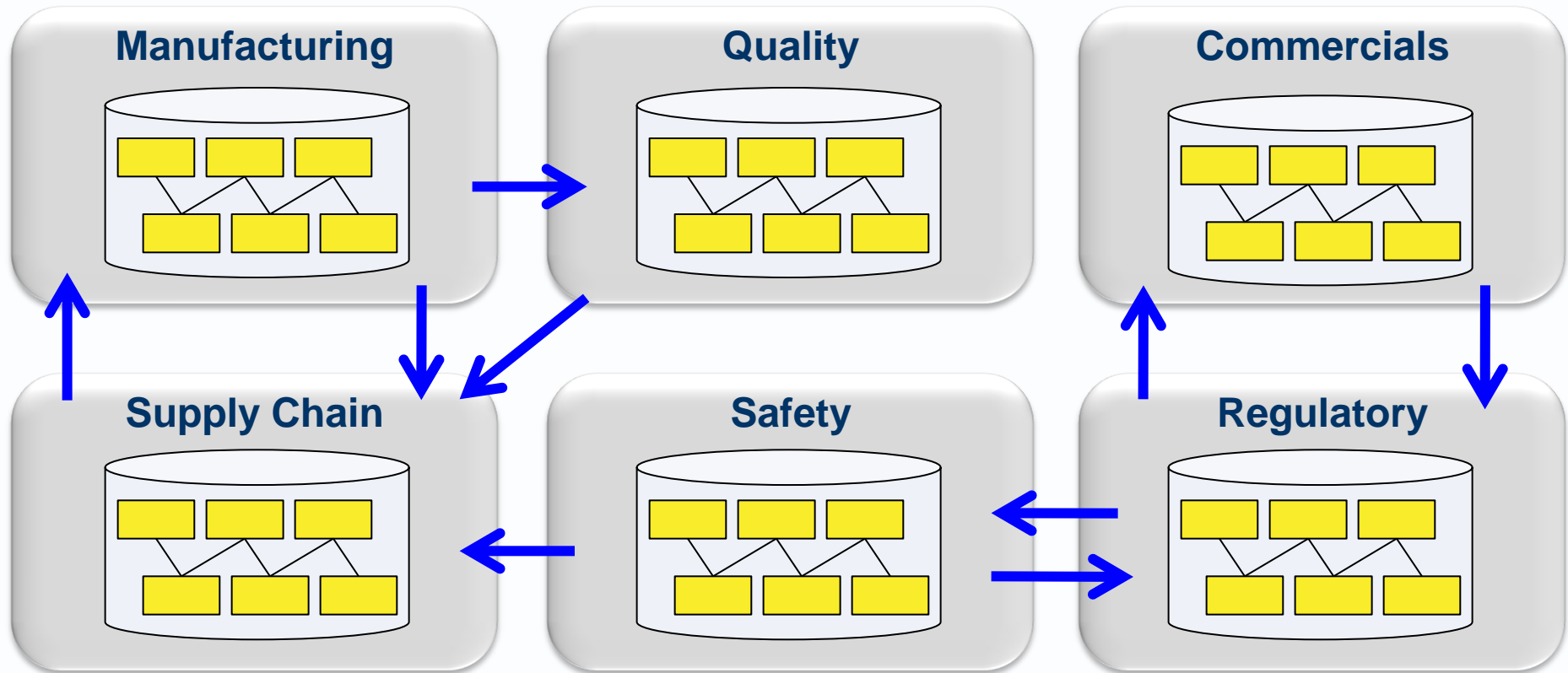
Business data in stores - duplication

► "Dear graham,

► Today's business users are accessing data from more sources than ever before which increases potential duplicate records, outdated information and allows for simple mistakes to slip through the cracks.

► More and more organisations are realising the negative impact that inaccurate and duplicate records cause within daily operations, and the expense of having to eliminate and manage this data.

email from Magic Software

Data Entity	CRM	ERP	Billing
Customer	x	x	x
Sale	x	x	x
Invoice			x

Avancier

Business data in stores – master data management

► "Addressing duplicate records efficiently requires a clear strategy and the tools to carry this out, such as:

1. Identify all internal systems that may contain duplicate records within the organisation;

2. Establish rules that define what a duplicate record is;

3. Define actions that occur for duplicate data;

4. List all duplicate records based on the criteria;

5. Build real-time or scheduled processes to eliminate duplicate records re-entering the systems."

email from Magic Software

Data Entity	CRM	ERP	Billing
Customer	x	x	x
Sale	x	x	x
Invoice			x

Data Entity	CRM	ERP	Billing
Customer	M	C	C
Sale	M(1)	C	M(2)
Invoice			M

Business data in stores - data dissemination matrix

► [An artifact] that tabulates data entities against data stores.

► It shows duplication of data between data stores.

► It is useful in analysis of change impacts, data mastering and security vulnerabilities.

► It may be used to define the master and copies of a data entity.

Common Entities Data stores	Customer	Product	Asset	Employee
CRM system.	Master			Copy
Call-center system.	Copy			
Contact-management system	Copy			
ERP system.		Master		
Order-processing system		Copy		
GL tracking			Copy	
Asset database			Master	
Timesheet				Copy
Expense Claim				Copy
Contract DB				Copy
Company Directory				Master

Avancier

Business data in stores - cluster analysis
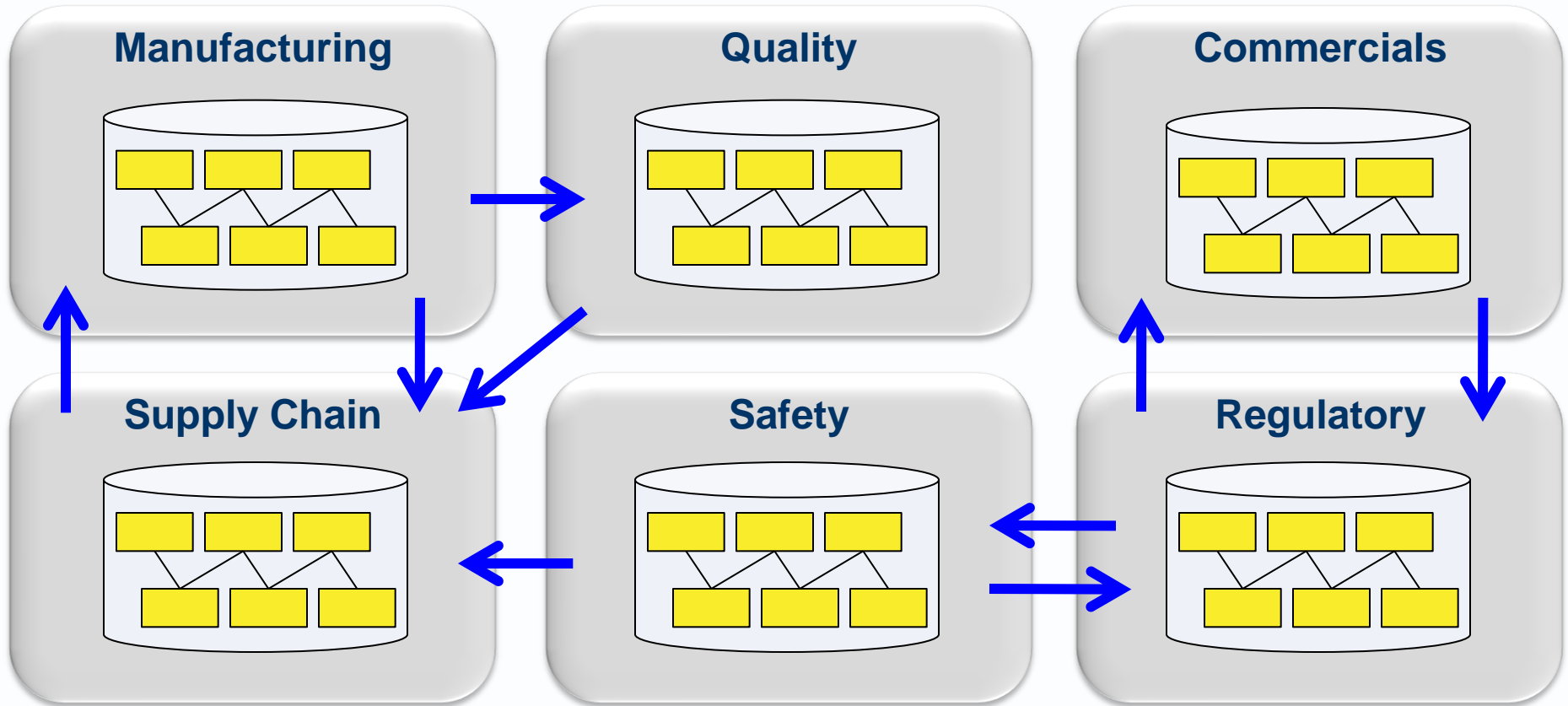
► Cluster analysis can be used to cluster

► data that is created by the same functions, and

► functions that create the same data.

Figure 12-1

© Minder Chen, 1997-2008

Enterprise A

► Identity data flows exchanged between functions

Business data flows - documentation

Data Flow (aka Function Communication) Diagram
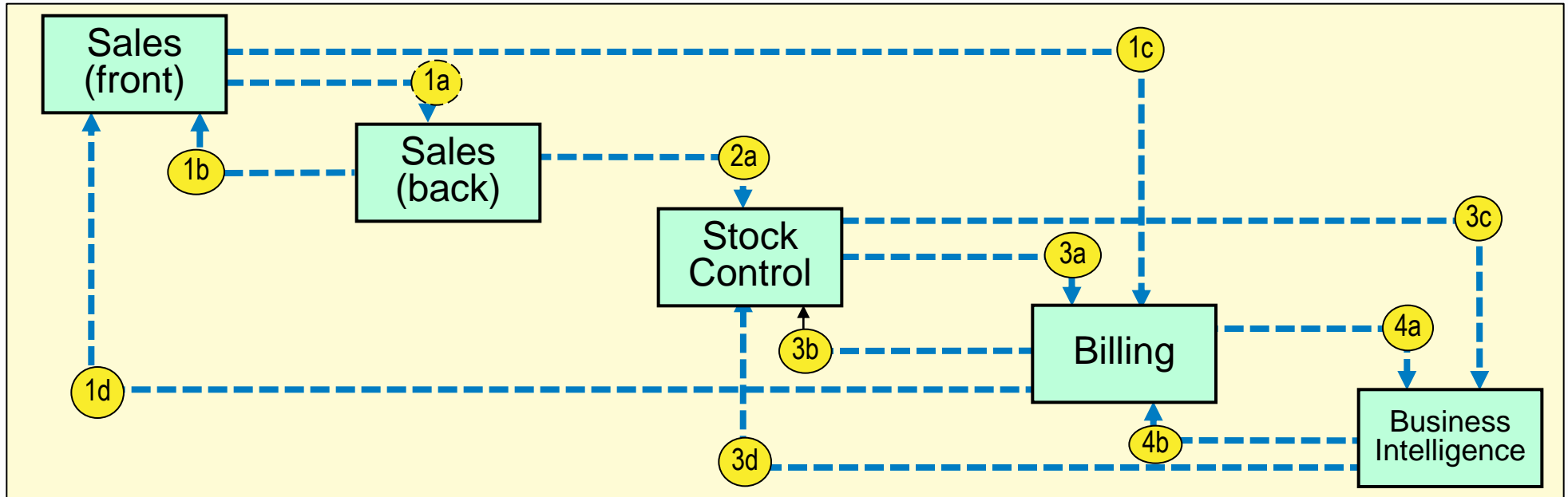
Data Flow (aka Interface) Catalogue

Id	Flow Name	Source	Destination	Content
1a	Order entry	Sale (front)	Sale (back)	Ref. 999
1b	Order accepted	Sale (back)	Sale (front)	Ref. 999
2a	Notification	Sale (back)	Stock Control	Ref. 999

The numbers problem

► In enterprise A, a large team took 18 months to document their many hundreds of business processes.

► In enterprise B, the business architects recorded 250 services, 250 capabilities, 75 processes at the first/top level and 1,000 processes at the fourth level of decomposition, 1,000 organisation units, 3,000 roles, and 5,000 employees.

► In enterprise C, they recorded 6,000 client-side apps and 1,000 server-side applications (including some that run on operating systems not supported by cloud service providers).

► A county council took 18 months to find and list 600 applications.

► A central government department found > 300 apps and > 500 data flows between them.

► Consider that central government department with > 300 apps and > 500 data flows between them.

► Suppose there are 30 data item types in each of the 500 data flows, that is **15,000 data item types.**

► But not all the data entered into the applications passes between applications.

► Suppose there 50 data entities in each of the 300 applications' data stores, that is 15,000 data entities and perhaps **50,000 attributes.**

► A larger enterprise will have more.

► The challenges of incorporating data structures from often very complex ERP systems into architectures are well known…

► **The sheer number of database tables** is enough to cause headaches, and coupled with customizations and multiple languages it **has a real impact on the effectiveness of your Enterprise Architecture.**

Corso and Silwood Technology
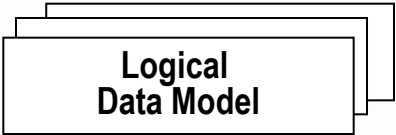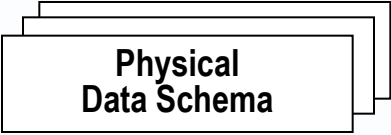
► How many tables SAP's ERP database schema?

► Usually, the enterprise data architect has to be selective in the data they study and record

► Leading bank is urgently seeking a proven EA to engage and lead IT projects including

- the Enterprise Information Architecture
- **information models** and flows,
- data dictionaries, data standards
- data quality standards and processes

► develop and maintain the logical Enterprise Information Architecture that enables seamless information interoperability of all Bank systems for efficiency and cost-effectiveness.

► eutopiaonline.com

Business data model?

	Data in at rest – in store	
Conceptual	Business Data Model	A data model that defines entities a business wants to remember, regardless of computing
Logical	Logical Data Model	
Physical	Physical Data Schema	
Real	Data Store	

Enterprise data architects must be selective

► Kernel data entities only

Sales and Marketing Domain

Campaign → Opportunity → Contact → Account / Contract
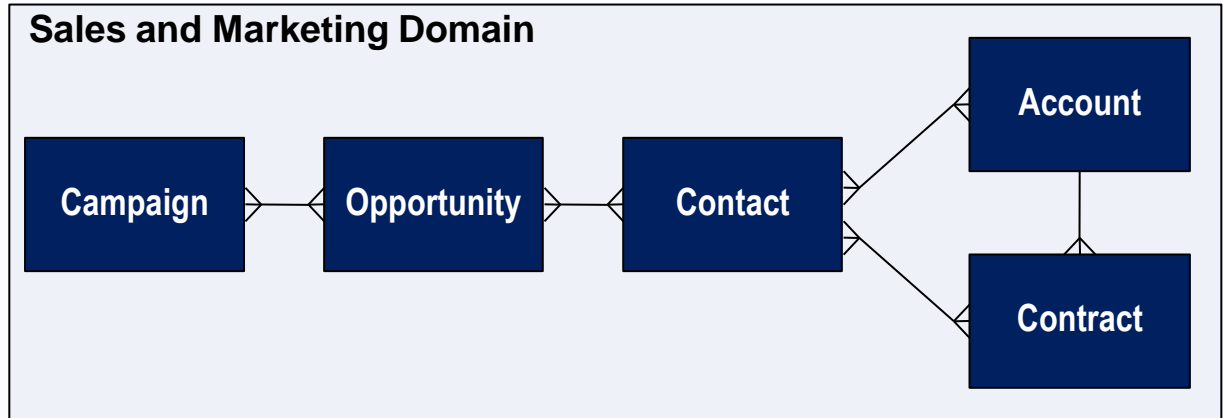
► Data entities in two or more data stores

City Council

Ward, Claimant, Directorate, Building, Business, Transaction, Worker, Member of Public, Enquiry

Avancier
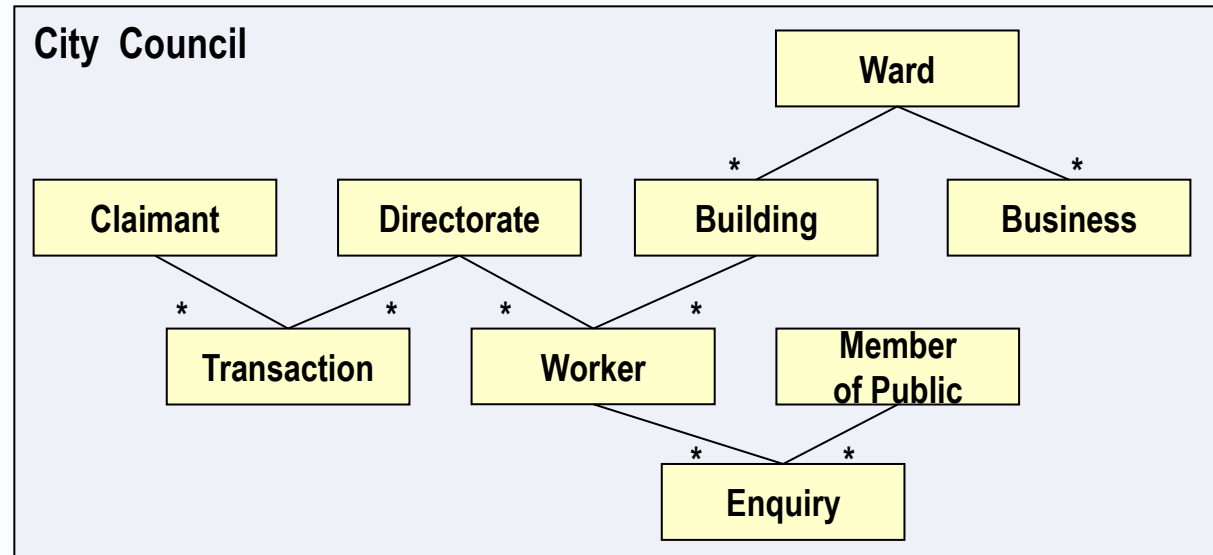
List **kernel data entities** under functions or more abstract headings

Question	Abstract classification	Data Entities
Why	Purposes	Goal, objective, requirement, measure
Who	Parties and people	Customer, supplier, organisation, employee
How	Processes	Transaction, event, order, payment
	Promotions	Campaign, advert, mailing.
What	Papers	Document, email, form
	Pounds and pennies	Account, budget, currency
	Properties	Maintained resource, office, vehicle, assets.
Where	Places	Area, invoice address, delivery address
	Pipes	Route, network
When	Points in time	Calendar, date, time

► Leading bank is urgently seeking a proven EA to engage and lead IT projects including

■ the Enterprise Information Architecture

■ information models and flows,

■ **data dictionaries,** data standards

■ data quality standards and processes

► develop and maintain the logical Enterprise Information Architecture that enables seamless information interoperability of all Bank systems for efficiency and cost-effectiveness.

► eutopiaonline.com

Business data flows between data stores
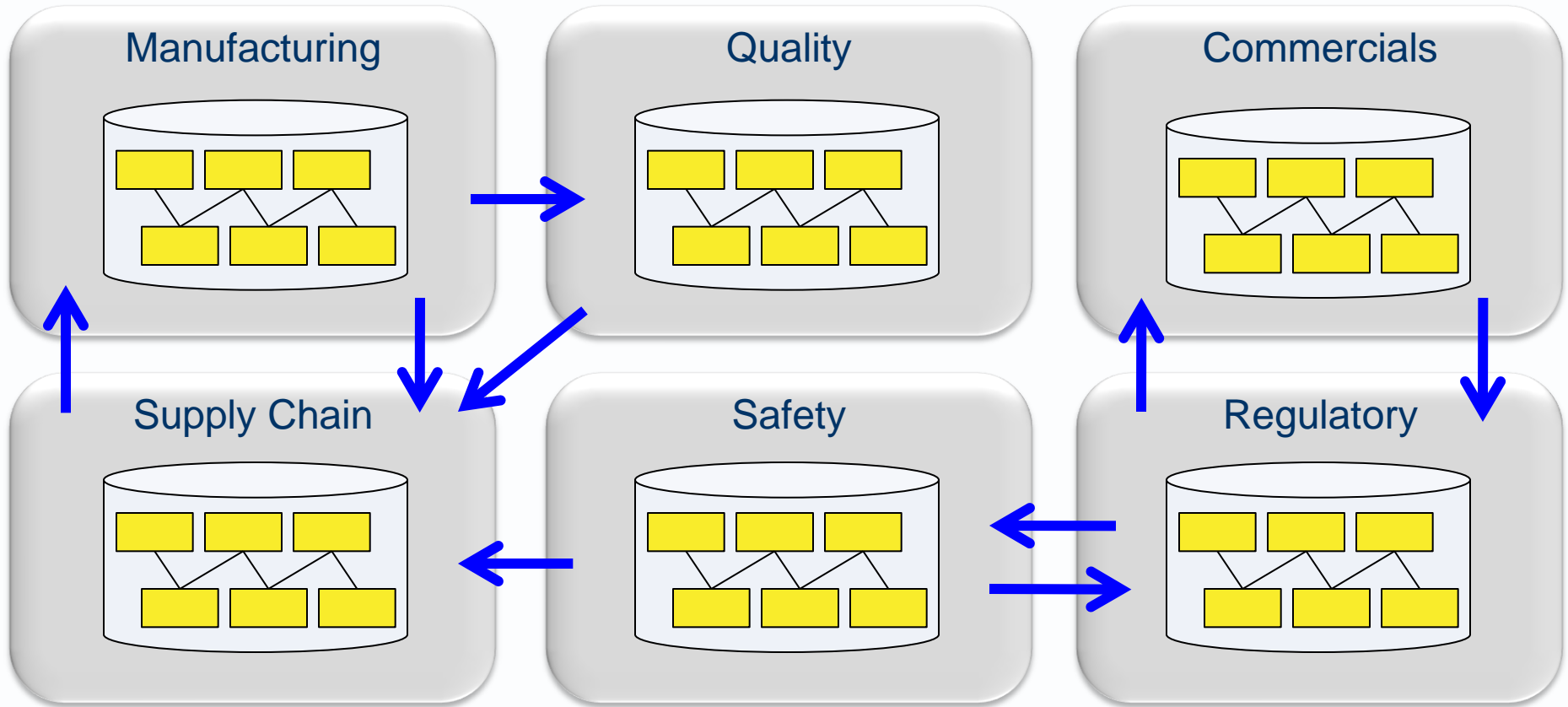
Enterprise-wide data definition?

	Data in at rest – in store	Data in motion in flow
Conceptual	*Relatively informal?* → **Business Data Model**	*A dictionary of standard or approved data types for communication* → **Canonical Data Model**
Logical	**Logical Data Model**	**Logical Data Flow Structure**
Physical	**Physical Data Schema**	**Physical Data Flow Structure**
Real	**Data Store**	**Data Flow**

Why do we need a data dictionary?

► Because terms like customer, policy, and profit mean different things to different people

► Shareholders want to know
- Q) How many vehicles has an Automaker sold this year?
- Q) At what profit per vehicle?

How many vehicles has Automaker sold this year?

► Procurement records the parts it buys, with supplier and price paid.

► Pre-assembly records parts assembled, with facts about their process.

► Assembly record **vehicles assembled**, with facts about the assembly process.

► Logistics record **vehicles shipped**, with facts about the shipment and receiving dealer.

► Automaker records a **vehicle sold** somewhere between assembly and shipment, but it may be returned as damaged or unsold.

► Each dealers records **vehicles purchased** for sale.

► Automaker's monthly sales report includes **vehicles sold by dealers**.

► The quarterly financial statement includes **unsold vehicles on dealer lots**.

► Automaker tracks when dealers lots are full, because they must give them costly incentives to help them shift the vehicles (and so, not slash orders).

1. Define entities and items in I/O data flows
2. Define data that must be remembered for future actions

Object	Description (edited from Salesforce.com)
Account	An individual account, which is an organization involved with your business (such as customers, competitors, and partners).
AccountContact	The role that a given Contact plays on an Account.
Role	A relationship between two Account objects, such as partnerships or subsidiaries.
AccountPartner	A sharing entry on an Account.
AccountShare	Associates a word or short phrase with an Account.
AccountTag	A User who is a member of an Account team.
AccountTeamMember	An additional phone number for a CallCenter.
AdditionalNumber	The AuthSession object represents an individual user session in your organization.
AuthSession	A link between two opportunities.

Data dictionary – a formal canonical data model

► [A standard] that provides the "one true definition" of data types and structures used in data flow structures.

► It define common data types and structures used in data flows
- Staff [Staff name, NiNo, DoB, StartDate]

► It may be defined at a logical level in some kind of data dictionary
► Or in an XML schema along with compression or codification of data
► So a message bus can transform all data flows into and out of the canonical data model form

Generic data format standards defined using XML

► A short tour of the non-technical industry efforts to create a common XML-based vocabulary for specified purposes and industries. PETE O'DELL "Silver Bullets"

► **1. Astronomy.** See http://fits.gsfc.nasa.gov.

► **2. Built environment, and infrastructure systems integration.** See www.obix.org.

► **3. Distribution/Commerce.** See www.rosettanet.org.

► **4. Education.** See schools interoperability framework.

► **5. Financial reporting. See www.xbrl.org.**

► **6. Financial research.** See www.rixml.org.

► **7. Food.** See www.mpxml.org.

► **8. Healthcare. See www.hl7.org.**

► **9. Information technology architecture.** (opengroup.org)

► **10. Instruments.** See www.nasa.gov

► **11. Insurance. See www. acord.org.**

► **12. Legal.** See www.legalxml.org.

► **13. Manufacturing.** See www.pslx.org. (no longer available)

► **14. News.** See www.iptc.org.

► **15. Oil and gas.** See www.pidx.org.

► **16. Publishing.** See www.oasis.org.

► **17. Real Estate.** See www.RETS.org.

► **18. Research.** See www.casrai.org.

► **19. Telecommunications.** See www.atis.org. + TMF

► FPML (financial products)
► FIXML (financial instruments)
► OASIS
  ■ Names? Addresses?
► Open Travel Alliance (OTA)
  ■ Cars, hotels, insurance, airports, currencies, countries
► GTFS (General Transit Feed Specification)
► TransXchange SIRI (Schedules)
► TRANSMODEL (EU public transport info.)
► Air travel – PNR passenger name record
► ARTS (association of retail, textile…)
► Open Geospatial Consortium (OGC)
► JISC – universities - HEDIIP

DAMA principles

DAMA's

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable
- **Data is an asset with unique properties**
- **The value of data can and should be expressed in economic terms**

- ► Consider the value of data to a business
- ► Address data qualities
- ► Define data using data
- ► Take a business-wide view
- ► Consider data over its life time

Data Management Requirements are Business Requirements
- **Managing data means managing the quality of data**
- **It takes Metadata to manage data**
- **It takes planning to manage data**
- **Data management requirements must drive Information Technology decisions**

Data Management depends on diverse skills
- **Data management is cross-functional**
- **Data management requires an enterprise perspective**
- **Data management must account for a range of perspectives**

Data Management is lifecycle management
- **Different types of data have different lifecycle characteristics**
- **Managing data includes managing the risks associated with data**

Data entity lifecycle diagram

► [An artifact] that shows life of a data entity in terms of

► **states** it passes through (from creation to deletion)

► **events** that trigger state transitions.
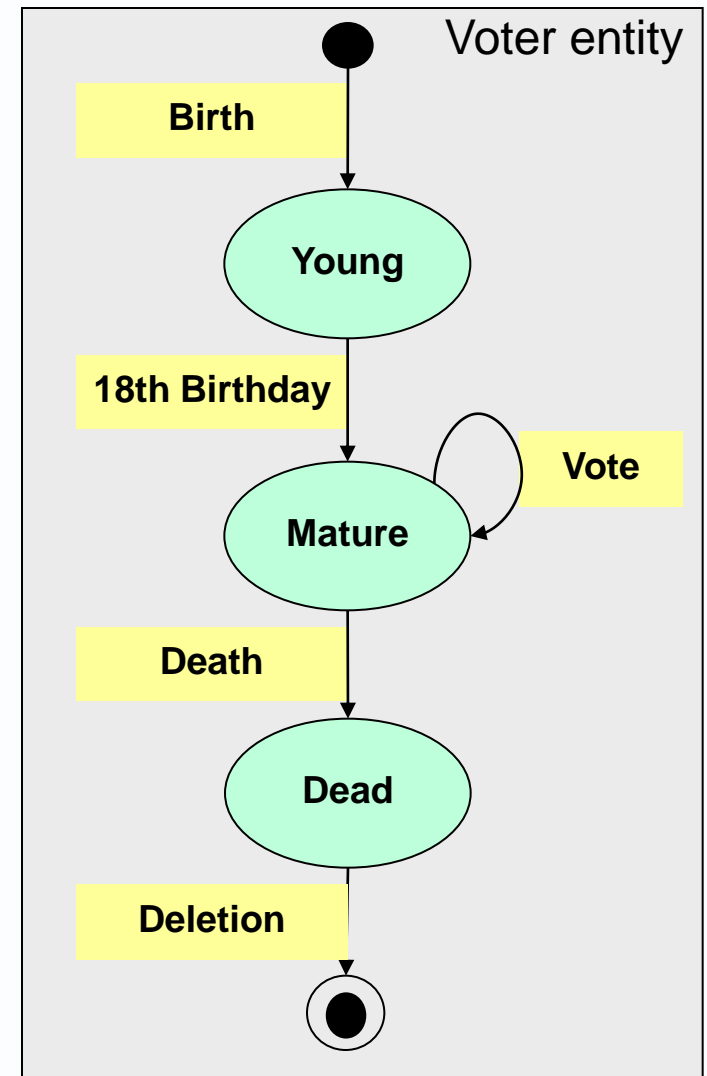
Voter entity

Birth

Young

18th Birthday

Vote

Mature

Death

Dead

Deletion

Avancier

► Leading bank is urgently seeking a proven EA to engage and lead IT projects including

- the Enterprise Information Architecture
- information models and flows,
- data dictionaries, data standards
- **data quality standards and processes**

► develop and maintain the logical Enterprise Information Architecture that enables seamless information interoperability of all Bank systems for efficiency and cost-effectiveness.

► eutopiaonline.com

DAMA principles

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable
- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms

► Consider the value of data to a business
► Address data qualities
► Define data using data
► Take a business-wide view
► Consider data over its life time

Data Management Requirements are Business Requirements
- Managing data means managing the quality of data
- It takes Metadata to manage data
- It takes planning to manage data
- Data management requirements must drive Information Technology decisions

Data Management depends on diverse skills
- Data management is cross-functional
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives

Data Management is lifecycle management
- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data

Data quality affects process quality

► Processes go wrong when data values are
- not consistent,
- not conformant to rules
- not correct
- not controlled

► when data types in different systems are
- miss-aligned (e.g. dates and addresses)
- wrongly thought to correspond (e.g. "profit")

► when data values are misinterpreted by end users who don't understand the meaning of the data type.

► The motivation and budget to clean the data will emerge if its lack of quality has a significant negative effect on its users and owners.

► So if you care about data quality:

- 1) don't store data that is not used,
- 2) make use of the data you store (a GDPR requirement)
- 3) document and report all negative effects of poor data quality.

► **Confidentiality**

> ► Secure from unauthorised access

► **Integrity**

> ► Consistent with other data
>
> ► Conformant to rules
>
> ► Correct - true
>
> ► Controlled – not open to change without authority

► **Availability**

> ► See design for NFRs

Data qualities – CIA-based security classification

► Tom Peltier rates the security level of a data item, structure or store as equal to the highest of the individual ratings for CIA.

	Confidentiality
	Impact of unauthorized use or disclosure
Customer	**High**
Order	Moderate
Product	Low

Data qualities – CIA-based software development rules

ID	Requirement Explanation	Factor			ID	Measure Explanation
		C	I	A		
AD002	Treat client code as unreliable. All client code (SPA / mobile) can be mutated by the user and must therefore be considered unreliable.	1	1	0	AD002.01	Validate input at least on the server. Validation on the client is for user experience only and offers no security whatsoever.
					AD002.02	Realize an input validation framework for common data types. Maximize the reuse of validations to prevent that marginal cases on specific actions are nevertheless accepted by the server and lead to security bugs
					AD002.03	Business rules are enforced on the server.Business rules on the client are for user experience only and do not offer any security.
AD006	Used external libraries / frameworks do not contain known vulnerabilities. Upon acceptance, all technical components used are validated to see if they have known vulnerabilities. If this is the case, the component must be replaced or updated	2	0	0	AD006.01	Check applied components against the CVE database. https://cve.mitre.org/cve/cve.html A collection of all publicly known security bugs in known systems / libraries.
					AD006.02	If the latest version is not used, check release notes for vulnerabilities that have occurred and resolved. Security bugs are often clearly stated in the release notes of components. Check if newer versions of the component fix important bugs and update the component if necessary.
					AD006.03	Apply the most recent version of the components and keep up with updates. Preferably always choose the most recent versions, although not always possible due to cross-dependency.
AD007	Sensitive data is not stored unnecessarily long. Minimize attack interface. Prevent leaks through sensitive information via debugging options / memory dumps and file access.	3	0	0	AD007.01	Sensitive data in the memory will be sorted / cleared as soon as it is no longer needed. This is to prevent memory dump leaks. Also take into account infrastructural and / or framework limitations.
					AD007.02	Sensitive data is removed from storage as quickly as possible when it is no longer needed. Avoid archived data that remains behind for a long time. Also take into account legal requirements with regard to retention periods.

AD?

AD refers the subset of measures - from Info Support – in this table
Does your organisation have such advice?

	Number of measures to implement
Architecture and Design	32
Authentication	19
Session Management	9
Access Control	13
Data Processing	17
Data at Rest	10
Data in Transit	7
Error Handling	6
Auditing	4
Distribution and Concurrency	3
Configuration	5
Total	125

Data ownership?

DAMA's

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable

- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms

► Consider the value of data to a business
► Address data qualities
► Define data using data
► Take a business-wide view
► Consider data over its life time

Data Management Requirements are Business Requirements

- Managing data means managing the quality of data
- It takes Metadata to manage data
- It takes planning to manage data
- Data management requirements must drive Information Technology decisions

Data Management depends on diverse skills

- Data management is cross-functional
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives

Data Management is lifecycle management

- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data

► Business owns business data; IT doesn't.

► **Data owner:** a business manager accountable for quality of data (e.g. employee pay grade, salary):
  - E.g.
    - the Registrar and student **data**;
    - the Treasurer and financial **data**;
    - the VP of Human Resources and employee **data**.

► Data owners appoint *data stewards*

Data stewards (nominated by data owners)

► **Data steward:** a person responsible for the management and fitness of *data* elements - both the content and metadata.

► **Data stewardship**: the management and oversight of an organization's data assets to help provide business users with high-quality data that is easily accessible in a consistent manner.

► Data owner/steward may (e.g.) decide which copy of data is master
  ■ Might also own the system that creates it and maintains it
  ■ But might not

► **Data Custodian** is a system administrator who has technical control over an information asset dataset.

GDPR adds

► Data Protection Officer

► Implies roles
  ■ Data controllers (in the organisation)
  ■ Data processors (authorised by controllers)

► Leading bank is urgently seeking a proven EA to engage and lead IT projects including

- the Enterprise Information Architecture
- information models and flows,
- data dictionaries, data standards
- data quality standards and processes

► develop and maintain the logical Enterprise Information Architecture that enables **seamless information interoperability**

► of all Bank systems for efficiency and cost-effectiveness.

► eutopiaonline.com

► "semantic interoperability

  ■ the ability of human and automated agents to coordinate their functioning based on a shared understanding of the data that flows among them

► is a major economic enabler."

► "Semantic interoperability problems drive up integration costs across industry."

EITAGlobal (www.eitaglobal.com)

Industry data format standardization efforts - using XML schema

► A short tour of the non-technical industry efforts to create a common XML-based vocabulary for specified purposes and industries. PETE O'DELL "Silver Bullets"

► **1. Astronomy.** See http://fits.gsfc.nasa.gov.

► **2. Built environment, and infrastructure systems integration.** See www.obix.org.

► **3. Distribution/Commerce.** See www.rosettanet.org.

► **4. Education.** See schools interoperability framework

► **5. Financial reporting. See www.xbrl.org.**

► **6. Financial research.** See www.rixml.org.

► **7. Food.** See www.mpxml.org.

► **8. Healthcare. See www.hl7.org.**

► **9. Information technology architecture.** (opengroup.org)

► **10. Instruments.** See www.nasa.gov

► **11. Insurance. See www. acord.org.**

► **12. Legal.** See www.legalxml.org.

► **13. Manufacturing.** See www.pslx.org. (watch out!)

► **14. News.** See www.iptc.org.

► **15. Oil and gas.** See www.pidx.org.

► **16. Publishing.** See www.oasis.org.

► **17. Real Estate.** See www.RETS.org.

► **18. Research.** See www.casrai.org.

► **19. Telecommunications.** See www.atis.org. + TMF

► FPML (financial products)
► FIXML (financial instruments)
► OASIS
  ■ Names? Addresses?
► Open Travel Alliance (OTA)
  ■ Cars, hotels, insurance, airports, currencies, countries
► GTFS (General Transit Feed Specification)
► TransXchange SIRI (Schedules)
► TRANSMODEL (EU public transport info.)
► Air travel – PNR passenger name record
► ARTS (association of retail, textile…)
► Open Geospatial Consortium (OGC)
► JISC – universities - HEDIIP

Implications for apps architecture?

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable
- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms

► Data management must drive IT decisions

Data Management Requirements are Business Requirements

- Managing data means managing the quality of data
- It takes Metadata to manage data
- It takes planning to manage data
- Data management requirements must drive Information Technology decisions

Data Management depends on diverse skills

- Data management is cross-functional
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives

Data Management is lifecycle management

- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data

Four apps architectures using a Canonical Data Model

4 line address   **3 line address**

Business		
Sender	Internet	4 to 3 Transform → Receiver

Sender ——————————→ 4 to 3 Transform → Receiver

Business   Internet   Business

---

6 line address

Sender → 4 to 6 Transform → 6 to 3 Transform → Receiver

CDM   CDM

Business   Internet   Business

Sender → 4 to 6 Transform → Store, Route & Deliver → 6 to 3 Transform → Receiver

CDM   CDM

App   Message Broker

Sender → 4 to 6 Transform → Store, Route & Deliver → 6 to 3 Transform → Receiver

App

CDM

Message Bus = Message Broker + CDM

App

Reusable Transformer

Tedious for point to point

Good for many to many

FYI: some data management standards

► Whitemarsh Short Paper Series on
  - Database Standards,
  - Database Design,
  - Data Interoperability and
  - Metadata Management
► Whitemarsh@Wiscorp.com

Exercise:

► Identity data flows exchange between functions
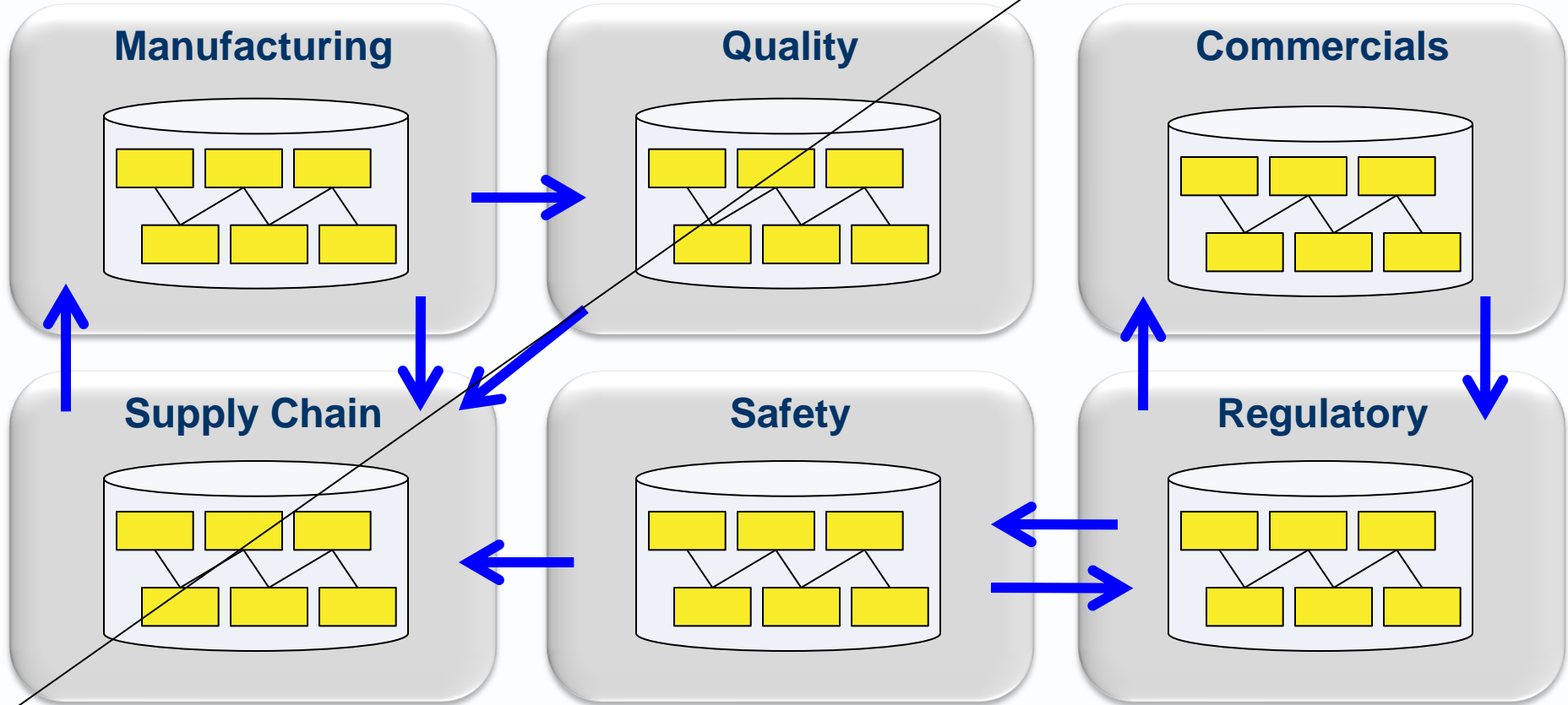
Exercise: data dissemination matrix

► Show which functions store which data entity

DATA ENTITY	Commercials	Supply Chain	Manufacturing	Quality	Safety	Regulatory

Exercise: Data flow catalogue (inter function)

DATA FLOW	Trigger	Sender	Receiver	Content