Deelnemerslijst
Data Management Fundamentals
17-19 november 2025

BEDRIJF	NAAM DEELNEMER		FUNCTIE
APG	Shanmukha	Masapu	business-analyst
Aegon Investment Management B.V.	André	Righolt	business-analyst
All Your BI	Ronald	van Gils	data architect
Bank Indonesia	Laras	Pramesti	data architect
European Investment Bank	Roberto	Alvarez	business-analyst
FMO (Nederlandse Financierings-Maatschappij voor Ontwikkelingslanden N.V.)	Stephanie	Allegrini	business-analyst
NAM	Duke	IGBUWE	it-manager

Evaluation Form
Data Management Fundamentals
November 17-19, 2025

Name: _____ Company: _____

What would be your overall grade on a 1 (worst) to 10 (best) scale (please tick):

			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1.	The workshop:		O	O	O	O	O	O	O	O	O	O
2.	The speaker:		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
		- Content:	O	O	O	O	O	O	O	O	O	O
		- Presentation skills:	O	O	O	O	O	O	O	O	O	O

3. Did the programme live up to your **expectations? [ ] Yes [ ] Partially [ ] |No, because**

_____

4. Would you recommend this workshop to colleagues or peers? **[ ] Yes [ ] No** (please explain)

_____

5. Which subjects did you miss or were not covered adequately?

_____

6. Which subjects were superfluous or took up too much time in your opinion?

_____

7. What additional comments can you offer?

_____

8. How would you grade the organisation and venue:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Overall organisation of the workshop	O	O	O	O	O	O	O	O	O	O
Quality classroom, screen and sound	O	O	O	O	O	O	O	O	O	O
Geographic location / accessibility	O	O	O	O	O	O	O	O	O	O

9. How did you come here? **[ ] Car [ ] Public transport [ ] Other** _____

10. Please leave your review/recommendation if you like, which we may post on our website.

_____

_____

_____

Name: _____

Job title: _____Company:_____

Please leave at our desk or mail to seminars@adeptevents.nl

Welcome

Adept Events

WHO WE ARE

AdeptEvents

DW&BI SUMMIT

BI-Platform.

RELEASE

Werner Schoots

Founder Adept Events

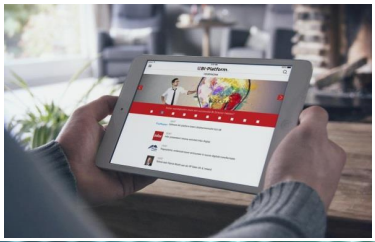
BI-Platform.

- Launched in 2008 as online spin-off from Database Magazine (DB/M)
- Topics: Business Intelligence, Data Warehousing, Analytics, Data Management

News	*Job board*	*Selected Whitepapers*	*Events*
Articles	*Blogs*	*Video interviews*	*Cases*

- We welcome your input: redactie@biplatform.nl

www.biplatform.nl & weekly newsletter
@BIplatform on & YouTube
Download the BI-Platform App
Join our LinkedIn Discussion Group



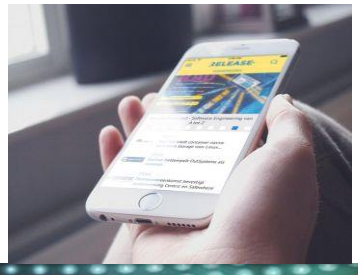
RELEASE.

- Launched in 1996 as Software Development spin-off from Database Magazine
- Topics: Software Engineering – Analysis, Design, Development, Testing and Deployment

News	*Job board*	*Selected Whitepapers*	*Events*
Articles	*Blogs*	*Video interviews*	*Cases*

- We welcome your input: redactie@release.nl

www.release.nl & weekly newsletter
@Release_nl on Twitter & YouTube
Download the Release App
Join our LinkedIn Discussion Group



SEMINARS

All seminars and workshops are organised twice a year

Alec Sharp	Business-oriented Data Modelling Masterclass Working with Business Processes Masterclass Concept Modelling for Business Analysts The Data-Process Connection (virtual half day session)
Panos Alexopoulos	Knowledge Graphs – pragmatic approach and best practices
Rick van der Lans	Ontwerpen van een Nieuwe Data Architectuur
Mathias Vercauteren	Data Governance Sprint
Nigel Turner	A Data Strategy for Becoming Data Driven Tackling Data Quality Problems (virtual half day session)
Chris Bradley / Winfried Etzel	Data Management Fundamentals
Lawrence Corr	Agile Data Warehouse Design & Dimensional Modeling
Christian Gijsels	Generatieve-AI in Business Analyse Cursus Sparx Enterprise Architect 16
Multiple speakers	*Data Warehousing & BI Summit – Yearly conference in March/April*

IN-HOUSE

All seminars and workshops can be organized in-company.
With local speakers and international speakers!

Please contact Werner Schoots

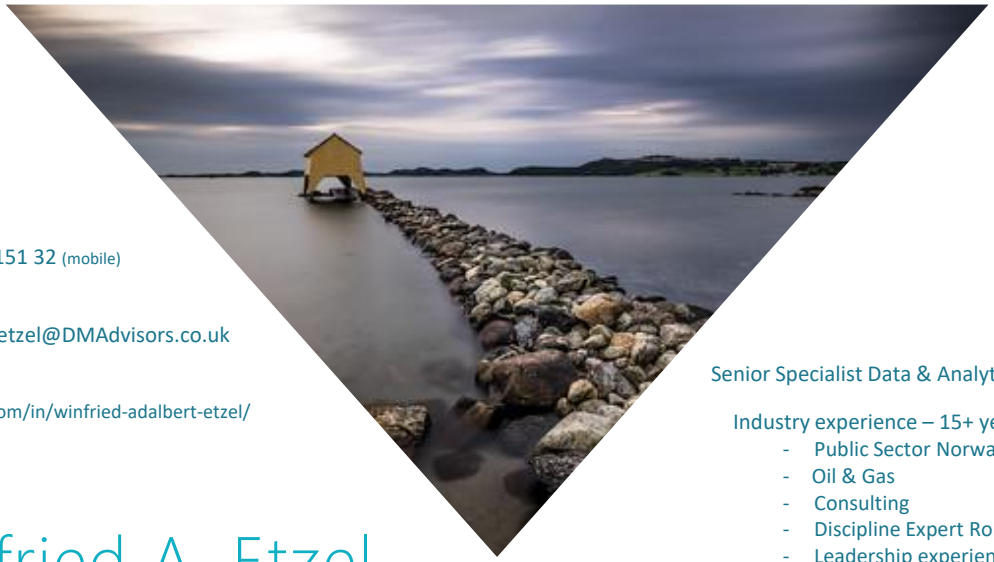
☎+31 (0)172 742680 ✉ seminars@adeptevents.nl

DMA
DATA MANAGEMENT ADVISORS

DMBoK2$_r$
CDMP Data Management Fundamentals

WINFRIED A. ETZEL CDMP Master

winfried.etzel@dmadvisors.co.uk



+47 942 151 32 (mobile)

winfried.etzel@DMAdvisors.co.uk

Linkedin.com/in/winfried-adalbert-etzel/

Winfried A. Etzel

INFORMATION MANAGEMENT STRATEGIST

Senior Specialist Data & Analytics - Equinor

Industry experience – 15+ years
- Public Sector Norway
- Oil & Gas
- Consulting
- Discipline Expert Roles
- Leadership experience in Data

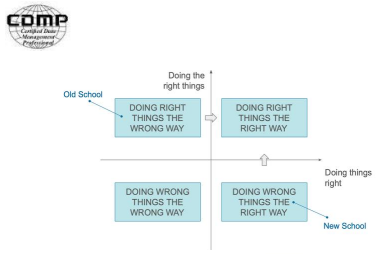
DMA
DATA MANAGEMENT ADVISORS

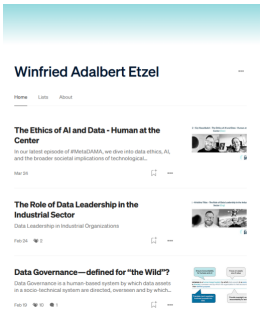
Winfried A. Etzel

Winfried Adalbert Etzel is a key figure in the **Nordic data community**, known for his strong contributions to professional forums and his commitment to promoting Data Governance, Data Strategy, and organizational design for data teams.

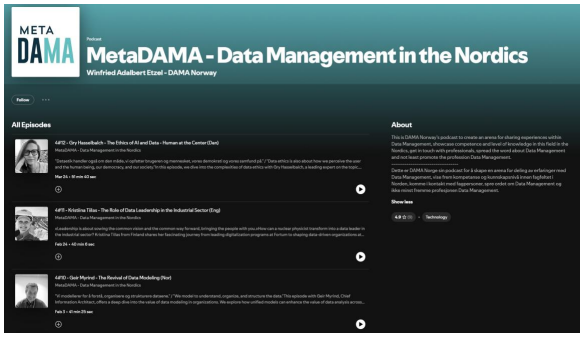
Winfried has dedicated his career to emphasizing the importance of information and data management, as well as a strategic and holistic approach to treating information and **data as a valuable asset** for organizations.

Through his work with the **#MetaDAMA podcast** and his voluntary engagement in the Nordic data community, he has been a driving force for innovation and has helped build a strong data community while promoting professional development across the Nordic region.





Winfried is currently working on his book under the working title **«Data Governance in the Wild»**, where he explores how Data Governance needs to change to adapt to distributed landscapes and the need for automation.



Recent Presentations

DAMA Webinar Series, February 2020: Becoming Data Centric? Data Strategy Part 1
Data Architecture Summit (DAS) 2019: Chicago USA, October 2019: "Essential competencies for Data Management Professionals" + "CDMP professional certification preparation"
DAMA Iceland, Reykjavik Iceland, September 2019: "Is data the "new oil"?
Data Management Forum, Bologna, Italy, August 2019: "Professional competencies for Data Governance"
Enterprise Data World 2019: Boston USA, March 2019: "Implementing a Data Strategy using the DMBoK2" + "CDMP professional certification preparation"
Riyadh Data Geeks, Saudi Arabia: December 2018: "Professional Data Management Certifications"
Lisbon, Portugal: October 2018: "Practical Data Governance"
Riyadh Data Geeks, Saudi Arabia: September 2018: "Data Integration & Interoperability"
Technology Transfer, Rome, June 2018 : "Data Modelling for Big Data"
MDM DG Europe, London, May 2018 : "Data Governance, Making it Real" + "Skills & Competencies for successful Data Governance"
EDW2018, San Diego, April 2018: "Data Governance, Making it Real"
Helsinki, Finland, March 2018: "Data Management Fundamentals"
Riyadh Data Geeks, Saudi Arabia: December 2017: "Practical Data Governance"
Rome, Italy, May 2017: "Advanced Data Modelling"
Lazio University, Rome, November 2017: Practical Data Governance
Data Management Lisbon Portugal: October 2017 "Practical Data Quality Management"
Harvard Business School, USA, September 2017: "Practical Data Governance"
University Warwick, September 2017: "Data Integration" MSc Information Management
Riyadh Data Geeks, Saudi Arabia: September 2017: "Data Quality Fundamentals"
Idera Seminars, London, August 2017: "Data Modelling for the people"
University West of England (UWE), June 2017: "Data Quality Management" MSc Information Management
MDM DG Europe, London, May 2017: "The role of a Conceptual Data Model in Data Governance" + "Developing Professional Data Management Competencies"
Rome, Italy, May 2017: "Data Governance & Data Management Fundamentals"
Helsinki Finland, May 2017 2017: "Data Management Fundamentals / Certification preparation"
Helsinki, Finland, March 2017: "Practical Data Governance"
Riyadh Data Geeks, Saudi Arabia: February 2017; "Data Governance Fundamentals"
Paris, France: December 2016; "IFP University, Petroleum Information Management Masters Degree | Information Management Fundamentals"
Lisbon, Portugal: November 2016; "Practical Data Governance"
Helsinki, Finland: November 2016; "Data Management Fundamentals / CDMP preparation"
Enterprise Data & Business Intelligence: London: November 2016; "What are the Vital Skills & Competencies necessary for Information Management Practitioners?"
Riyadh Data Geeks, Saudi Arabia: October 2016; "Data Management Disciplines of the DMBoK"
Data Management LATAM Sau Paulo: September 2016; "Data Modelling is NOT just for DBMS's"
DAMA Sydney: August 2016 Sydney; "Data Quality, The Impact of Dirty Data?"
DAMA Melbourne: August 2016 Melbourne; "Data Quality, Are You Feeling Lucky?"
DAMA Canberra: August 2016 Canberra; "Data Quality, Are You Feeling Lucky?"
Australian Computer Society (ACS): August 2016 Canberra; "Data Quality Matters For Machines Too"
MDM DG Europe: May 2016 London; "Data Governance by Stealth"

Webinar: May 2016; "Data Modelling is not JUST for DBMS design"
Enterprise Data World: April 2016 San Diego; "DAMA CDMP Workshop"
DAMA Webinar: April 2016; "Data Integration & Interoperability" Disciplines of the DAMA DMBoK"
DAMA Webinar: March 2016; "Document & Content Management" Disciplines of the DAMA DMBoK"
DAMA Webinar: February2016; "Data Operations Management" Disciplines of the DAMA DMBoK"
Oil & Gas Data Management Conference: February 2016, London; "Developing An Information Strategy To Align With In-Flight Business Programs"
DAMA Webinar: January 2016; "Data Governance" Disciplines of the DAMA DMBoK"
DAMA Webinar: December 2015; "Information Lifecycle Management" Disciplines of the DAMA DMBoK"
DAMA Webinar: November 2015; "MetaData Management" Disciplines of the DAMA DMBoK"
Enterprise Data & BI Europe (IRM): November 2015, London; "Is the Data Asset really different#" & "CDMP Examination Preparation" & "Data Management Room 101"
DAMA Webinar: October 2015; "Data Risk & Security" Disciplines of the DAMA DMBoK"
DAMA Webinar: September 2015; "Data Warehousing & Business Intelligence" Disciplines of the DAMA DMBoK"
DAMA Webinar: August 2015; "Data Quality Management" Disciplines of the DAMA DMBoK"
BCS & DAMA Seminar: June 2015; "Is the Data Asset really different#?"
DAMA Webinar: June 2015; "Data Modelling" Disciplines of the DAMA DMBoK"
PRISME Pharmaceutical Congress: May 2015, Basel, CH; "Building & exploiting a Pharmaceutical Industry consensus data model"
MDM DG Europe (IRM): May 2015, London; "CDMP Examination Preparation" & "Data Governance By Stealth#, Can you 'sell' Data Governance if the stakeholders don't get it?"
DAMA Webinar: April 2015; "Master & Reference Data Management" Disciplines of the DMBoK"
Enterprise Data World: April 2015, Washington DC USA; "Data Modelling For The Business" and "Evaluating Information Management Tools **DAMA Webinar:** February 2015; "An Introduction to the Information Disciplines of the DMBoK"
Dataversity Webinar: February 2015; "How to successfully introduce Master & Reference data management"
Petroleum Information Management Summit 2015: February 2015, Berlin, Germany, "How to succeed with MDM and Data Governance"
Enterprise Data & Business Intelligence 2014: (IRM), November 2014, London, UK "Data Modelling 101"
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 Modelling 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 a 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"

Recent Publications

Book:	"DAMA Data Management Body Of Knowledge 2.0" | Technics Publishing | July 2017 | ISBN 978-1-63462-234-9
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:	**Data is NOT the new Oil January 2018**
Article:	Back to the future for Data management? September 2015
Article:	Is the "Data Asset" really different? July 2015
Article:	A visit to the vet & a BA flight reminded me about Data Governance; June 2015
White Paper:	"Information is at the heart of ALL Architecture disciplines",; March 2014
White Paper:	"Are you ready for Big Data ?", November 2013
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

Data is NOT the "new" Oil ...

The "Data Asset" is different

I presented a seminar at an event called "Data, the vital organisation enabler" entitled *"Information is at the Heart of ALL of the Business"*. During this I raised the question, "Is the data asset really that different from other assets?"

We hear copious amounts said that Data is an asset, it's got to be managed, few people in the business understands us and so on.

Don't get me wrong, I'm not trying to cast any doubt on the importance of data as an asset, but I wanted to raise the level of debate from a subliminal nod to a conscious examination of the characteristics of different "assets" and to compare them with the 'Data asset'.

Is Data the new Oil?

But before I go on much further, I want to challenge the often-heard statement "Data is the new Oil". Having studied Chemical Engineering at University and worked for 30 years in Data Management, much of this in the Oil industry (Aramco, BP, ExxonMobil, Shell, Statoil, Total to name a few), I now really believe that Data is NOT the new Oil.

Sure, the catchy phrase may be trying to indicate that we have a data economy vs an Oil economy, or maybe it's that Data is the lubricant of modern business, or maybe that there are Data "mega firms" that control much of our lives like the Oil mega-firms. However, while it is likeable, the analogy risks spreading misunderstanding on all counts.

We'll see later in this paper a comparison between Data and other Assets, but first let's drill in deeper to Oil 🛢 and look at different characteristics of this Asset.

Supply: There is a fixed amount of oil on the Planet. We also have Oil Shale, Oil Sands, Hydraulically Fractured Gas, Algae and so on. So yes, there is still plenty remaining, and we certainly haven't discovered all of it yet. But without a doubt its supply is finite. However, data is virtually limitless. Its sources and provision are plentiful. So, thinking of Data "supply" it's more like sea-water than oil. In fact, there is so much of it that our main worry is how to sensibly get value from it rather than how to find it.

Copyable: Whilst we may be able to synthesise certain lubricants (I ran my race car with synthetic oil) Oil is not readily copyable. However, as we know Data can be copied, and copied, and copied (with all of the resultant Governance issues that throws up), however the digital fidelity of the copy remains constant.

Use depletes it: As a petrol head both my race-car and hot road car have the unfortunate tendency to drink fuel whenever I put my foot down. Not great mile per gallon, but with all that power I get plenty of smiles per gallon. So, when the fuel is used – it's gone. Similarly, if we have Oil fields in the North Sea, the licence owner will have control of the oil there, but once it's used up, it's gone.
Data however can used and used and used again. The use of data doesn't deplete it. Also, the same data can be used by different people for different purposes (and without good Data Management also leads to Governance issues).

Value: It is most certainly usual practice to give a value to oil (the $70 barrel for example). The value of Oil is known on the commodities markets and has a real, identifiable asset value.
The practice of giving a real monetary value to Data is relatively immature (albeit growing). A small number of companies have a P&L entry for "value of Data" and this trend is increasing. It's in a similar position to where "Value of Goodwill" was on Balance sheets 20+ years ago, yet now its common practice.

Process to Yield Value: Although as noted in "Value" above, Oil has an established value, it is true that both data and Oil need to be processed in order to yield something of greater value than the raw base asset itself. With data, we have the well-established data value chain (Data > Information > Knowledge > Wisdom (insight). With Oil the products that emerge from processing can include Petrol, Diesel, Asphalt, Tyres, Antiseptics, Refrigerant or Plastic.

http://publications.dmadvisors.co.uk/articles/data-is-not-the-new-oil/

Training, Mentoring, and Executive Workshops

A DAMAI Registered Education Provider (REP)

We offer training courses for practitioners and executives. Custom-built, training & awareness seminars can also be delivered.

Building a business focused Data Strategy - 3-day course showing the components of a business focused data strategy, how to gain input and but-in, how to develop realistic implementation roadmaps and transition plans, and what it really means to be "data centric".

Data Management Fundamentals – **5-day** intermediate course covering all of the disciplines of Information Management as defined in the DAMA Body of Knowledge (DMBoK) & the changes in DMBoK 2.0. Core aspects & methods of the Information disciplines are explored.

Introduction to Information Management – **1-day** high level overview course, introducing the major disciplines of Information Management, why it is critical for business today, and the core subjects within the Information Management topic.

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. & how data models should be used for Business improvement & understanding.

Advanced Data Modeling (including Data Modelling for Big Data) – **3-day** advanced course for students with existing data modelling experience to understand the human centric aspects and techniques of data modelling to enable them to build quality models that meet business needs. This also covers major data model patterns & common problem solving. The course also show the applicability of Data Modelling in Big Data, NoSQL, and other non-relational environments.

IM Fundamentals & Practitioner Courses – A series of **1-day** *(foundation)* and **2–3-day** *(practitioner)* classes to provide practitioners a solid background in a specific Information Management topics. The practitioner workshops explore further detail on the implementation aspects of the Information Management discipline & also cover the CDMP specialist exam syllabus, providing preparation for students seeking to take the specialist CDMP exam.

- Data Modelling Foundation (*1 day only*)
- Data Governance & Stewardship
- Master & Reference Data Management
- Data Quality Management
- Data Warehouse & Business Intelligence
- Data Integration & Interoperability
- Metadata Management

Executive Workshops – **½ day** 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.

Professional Certification – **3-day** and **5-day** workshop "exam cram" designed to help attendees prepare for sitting professional certifications including DAMA CDMP examinations.

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.

DMA DATA MANAGEMENT ADVISORS

DMBoK2 and CDMP® Preparation

Live and On-Demand Classes

Data Management Fundamentals

https://www.dataversity.net/dmbok-and-cdmp-preparation-learning-plan/

Data Governance

https://training.dataversity.net/learning-paths/dgs0-dmbok-and-cdmp-preparation-data-governance-specialist-learning-plan

Data Modelling

https://training.dataversity.net/learning-paths/dms0-dmbok-and-cdmp-preparation-data-modeling-specialist-learning-plan

Data Quality Management

https://training.dataversity.net/learning-paths/dqs0-dmbok-and-cdmp-preparation-data-quality-specialist-learning-plan

Master & Reference Data Management

https://www.dataversity.net/

Metadata Management

https://www.dataversity.net/

- CDMP Online learning program
- Approved by DAMA-I
- Part of DATAVERSITY training center
- Based on DMBoK CDMP syllabus

20% discount code for DAMA UK **"dmadvisors"**

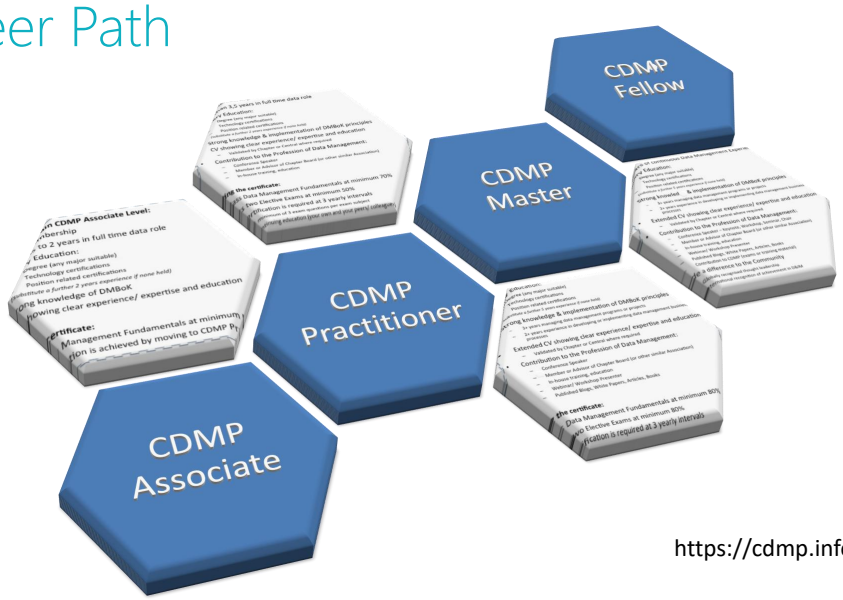
DMA DATA MANAGEMENT ADVISORS

Please Introduce yourself

1. Where you work & for how long
2. Information Management experience
3. Key learning objective(s) from this course (e.g. are you seeking to take the CDMP DMF exam later)
4. One **interesting <u>non-work</u>** fact about you

DMA
DATA MANAGEMENT ADVISORS

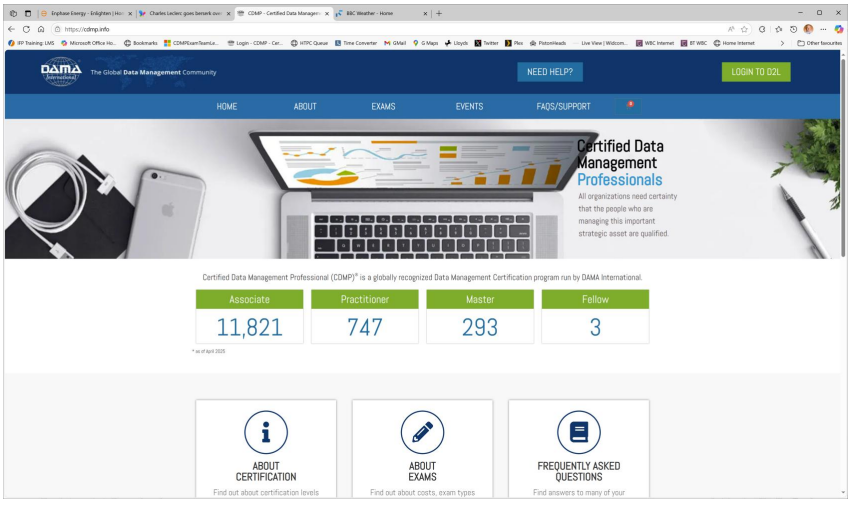
CDMP – Career Path



https://cdmp.info/

DMA
DATA MANAGEMENT ADVISORS

CDMP Portal: https://cdmp.info



DMA
DATA MANAGEMENT ADVISORS

Value of the CDMP?

Individual Benefits:

- Professional Development Milestones
- Continuous Professional Growth
- Internationally Recognized Measurement of Accomplishment
- Greater credibility and recognition in the marketplace
- May be a requirement in some areas

Organizational Benefits:

- A benchmark for assessing Data Management practices
- Increased trust in your Data Management team
- Employee development
- Demonstration of commitment to quality standards
- Raise awareness of the importance of Data Management disciplines to the organization

DMA
DATA MANAGEMENT ADVISORS

CDMP Associate

- 1 exam – **Data Management Fundamentals**
- Based upon DMBoK2 (see breakdown for %)
- Associate pass mark 60%
- ***Immediate*** award of CDMP Associate for passing DM Fundamentals with 60%+ pass

Examination

- 100 questions
- Multiple choice .. Only one answer per question
- 90 minutes (+20 mins where English is 2nd Language (ESL))
- 60% to pass

DMA
DATA MANAGEMENT ADVISORS

CDMP Practitioner

3 exams

- **Data Management Fundamentals** (70% required to pass)
- **2 Specialist exams** (70% required to pass)

Specialist Exams (each)

- 100 questions
- Multiple choice .. Only one answer per question
- 90 minutes (+20 mins where English is 2nd Language)
- 70% to pass

Specialist Exams

- Data Governance
- Data Modelling
- Data Quality
- Metadata
- Data Integration
- DW / BI
- Master & Reference Data
- *Data Architecture Management (on hold for DMBoK 3)*

DMA DATA MANAGEMENT ADVISORS

CDMP Master

3 exams

- Minimum 10 years Data Management experience (Evidence of experience reviewed by CDMP Fellows)
- **Data Management Fundamentals** (80% required to pass)
- **2 Specialist exams** (80% required to pass)

Specialist Exams (each)

- 100 questions
- Multiple choice .. Only one answer per question
- 90 minutes (+20 mins where English is 2nd Language)
- 80% to pass

Specialist Exams

- Data Governance
- Data Modelling
- Data Quality
- Metadata
- Data Integration
- DW / BI
- Master & Reference Data
- *Data Architecture Management (on hold for DMBoK 3)*

DMA DATA MANAGEMENT ADVISORS

CDMP Levels

		Associate	Practitioner	Master	Fellow
	DAMA Membership	Central	Central	Central	Central
Practical Experience	**Guideline**[2]	6 months > 5 years	2 years – 10 years	10 years ***minimum***[2]	Over 25 years[2]
Practical Experience	**Maximum**	N/A[1]	N/A[1]	N/A[1]	N/A[1]
	Exams	1 DM Fundamentals	3 DM Fundamentals + 2 specialist	3 DM Fundamentals + 2 specialist	Globally recognised & respected thought leadership. Well published. Significant contribution to Data Management Profession. CDMP Master. Contribution to CDMP & DMBoK. By nomination.
	Allowable substitutions[3]	No	No	No	
	Pass Mark	60%	70%	80%	
	Route	Register & take exam	Register & take exams	Register & take exams. Evidence experience (reviewed by CDMP Fellows)	

1. Individuals are welcome to remain at any level without pressure to go upward. You can decide what level you want to be and if you wish to proceed further.
2. Industry experience is a *guideline* for taking the certification. For Master, a minimum level of experience is required.
3. Substitute exams ceased from 1st August 2020.

DMA DATA MANAGEMENT ADVISORS

CDMP Exam Prices

Item	Member
All CDMP exams (taken at conferences, training events, chapters etc..)	$300
Online proctoring fee	$11
CDMP Renewal	$100
Exam re-take	$200
All exams are taken online only (classroom proctored or online proctor) *Online or chapter-led proctoring requires full payment up front* *Please watch out at various international DAMA-I endorsed conferences for exam proctoring and preparation workshops with Pay If You Pass offers.*	

https://cdmp.info/
support@cdmp.info
cdmp@DAMA.org

DMA DATA MANAGEMENT ADVISORS

Question bank

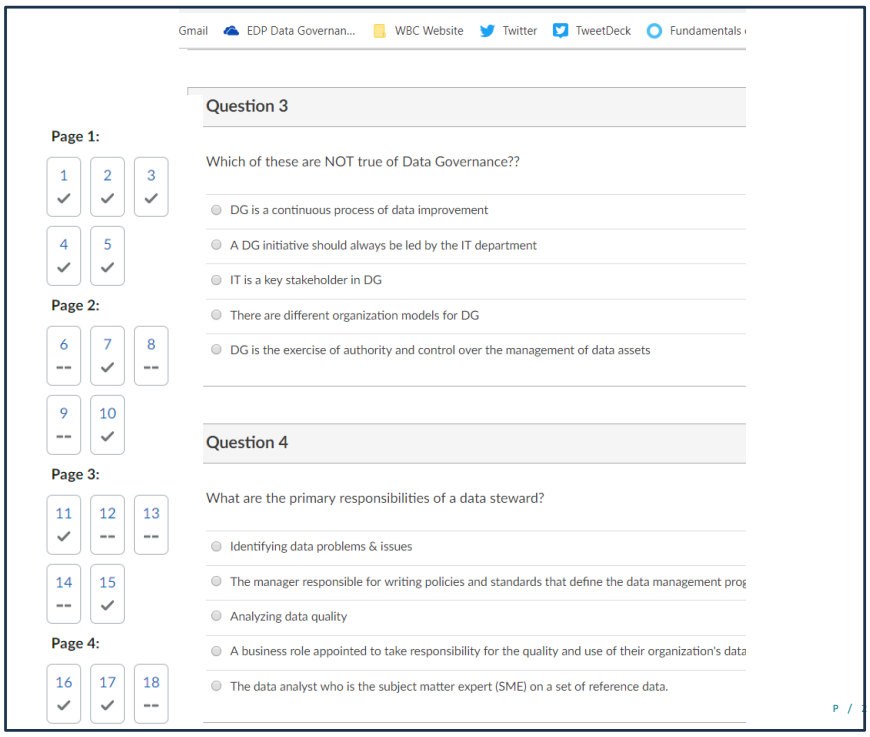
Difficulty Level	Knowledge Tested	% of q's in exam pool
Associate level difficulty	DMBoK2 content and understanding	60%
Practitioner level difficulty	Practical Application of DMBoK2	20%
Master level difficulty	Knowledge expected of a professional with 10+ years practical experience in the topic*	20%

*unlikely to be found in DMBoK2

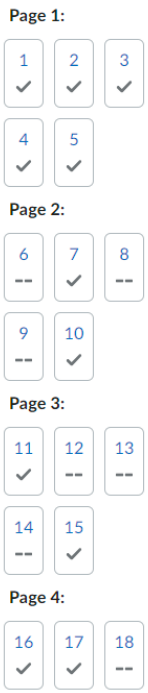
Module breakdown: DM Fundamentals

Data Management Process	Ch 1 + 15, 16, 17	2%
Big Data	Ch 14	2%
Data Architecture & L/C management	Ch 4	6%
Document, Records & Content Management	Ch 9	4-5%
Data Ethics	Ch 2	2%
Data Governance	Ch 3	11%
Data Integration & Interoperability	Ch 8	6-7%
Master & Reference Data Management	Ch 10	11%
Data Modelling & Design	Ch 5	11%
Data Quality	Ch 13	11%
Data Security	Ch 7	6%
Data Storage & Operations	Ch 6	4-5%
Data Warehousing & Business Intelligence	Ch 11	11%
Metadata Management	Ch 12	11%

Exam Platform



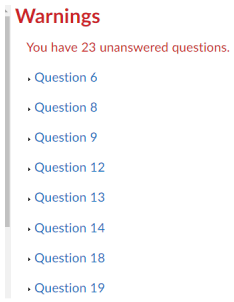
DMA
DATA MANAGEMENT ADVISORS



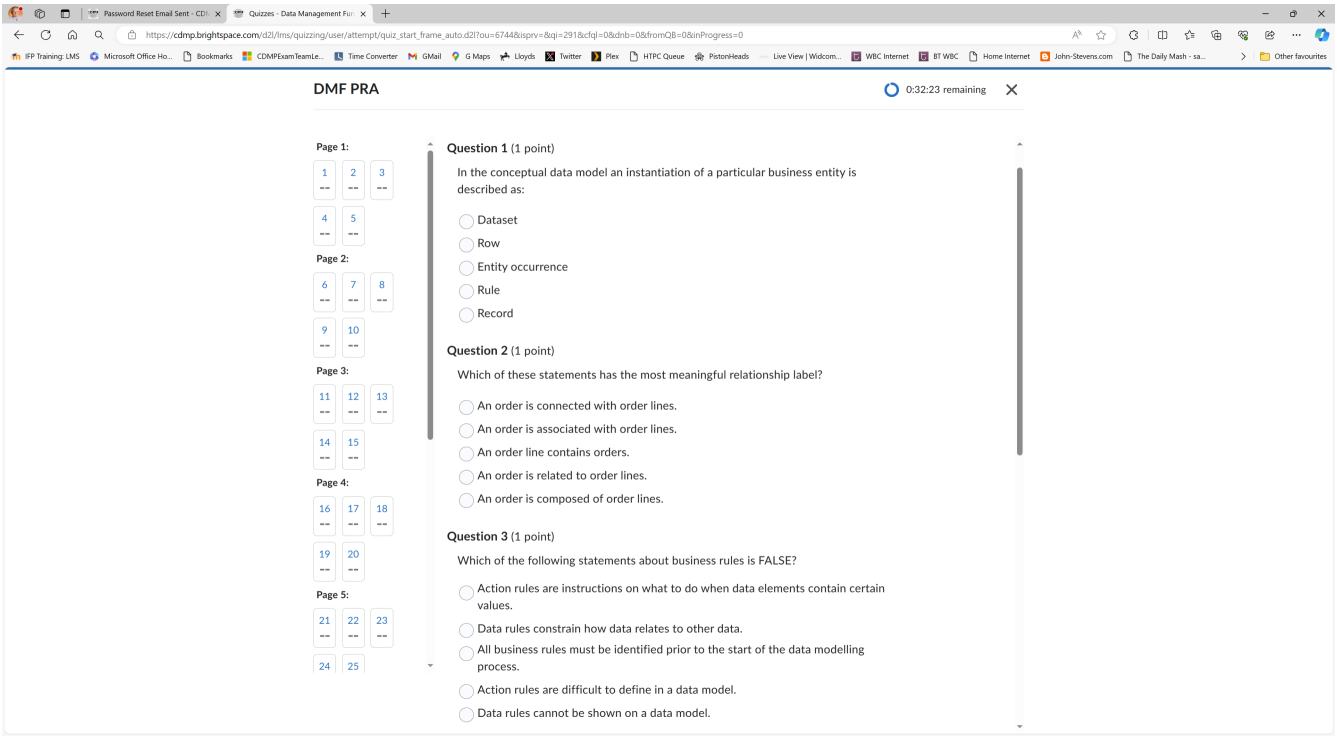
Exam Platform

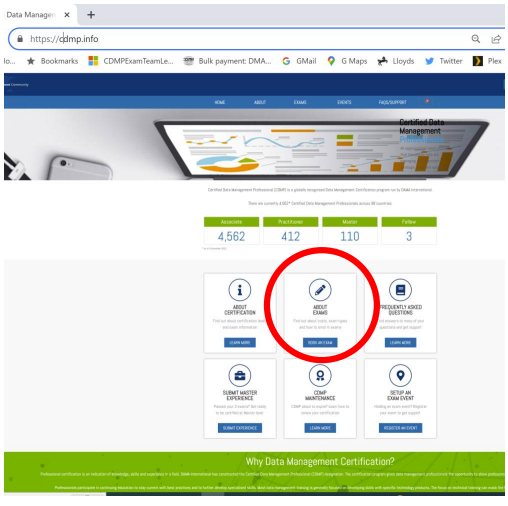
- Answered & Unanswered questions clearly indicated
- At submission, warning of unanswered questions listed ….
- …. click on these & answer before confirming submission
- As you proceed, note / flag any question to come back to

But always give an answer NOW



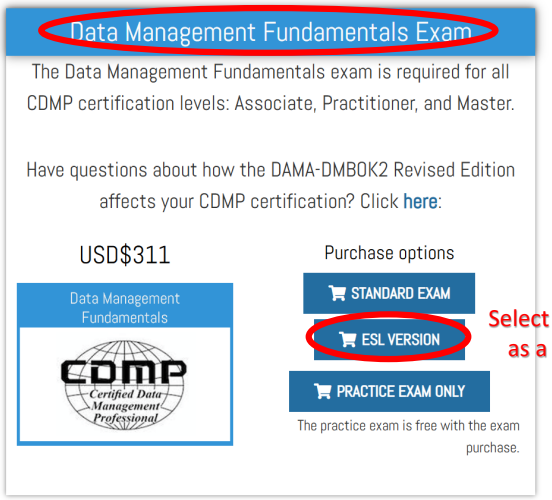
DMA
DATA MANAGEMENT ADVISORS





Purchase exams

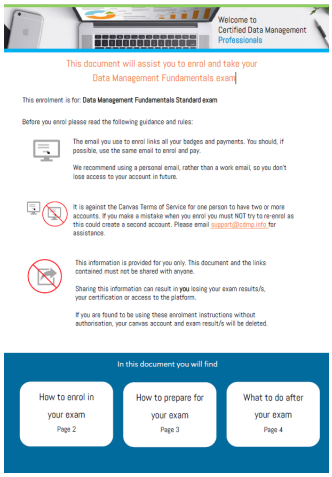
Purchase Data Management Fundamentals exam

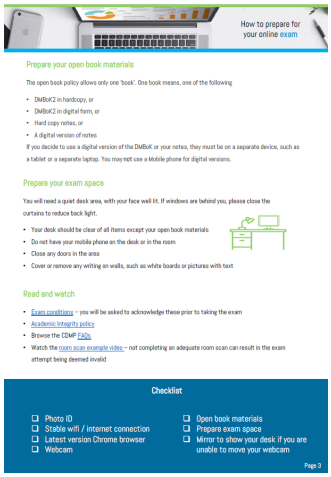


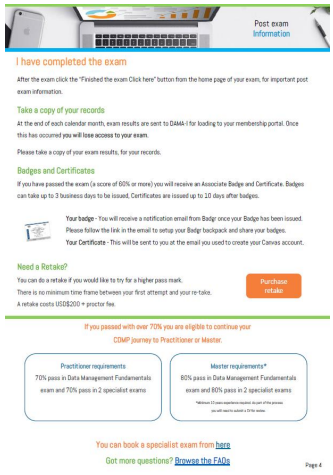
Select version (e.g. English as a Second Language)

DMA
DATA MANAGEMENT ADVISORS

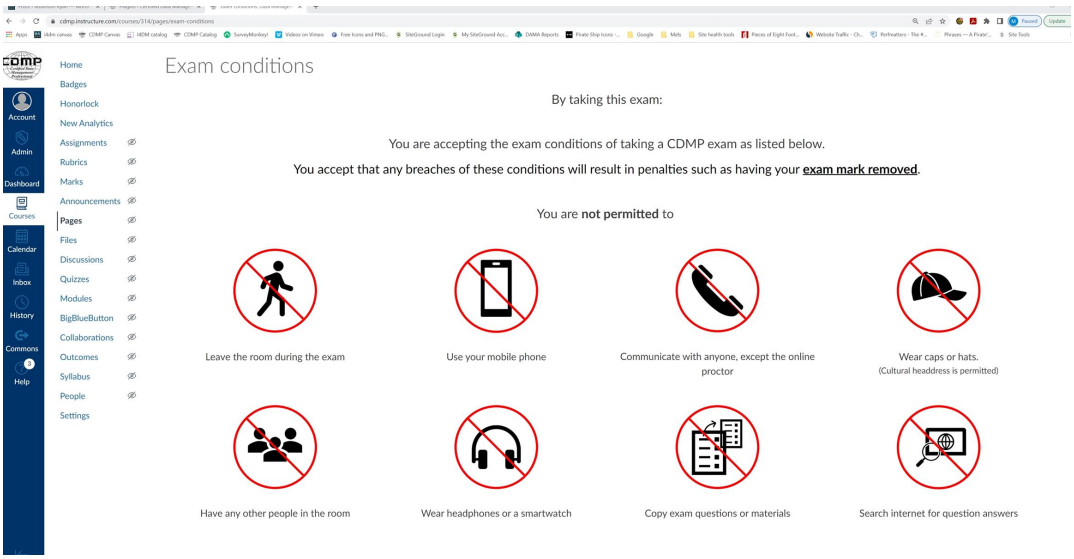
After purchase, follow the emailed instructions to register an account on the examination platform



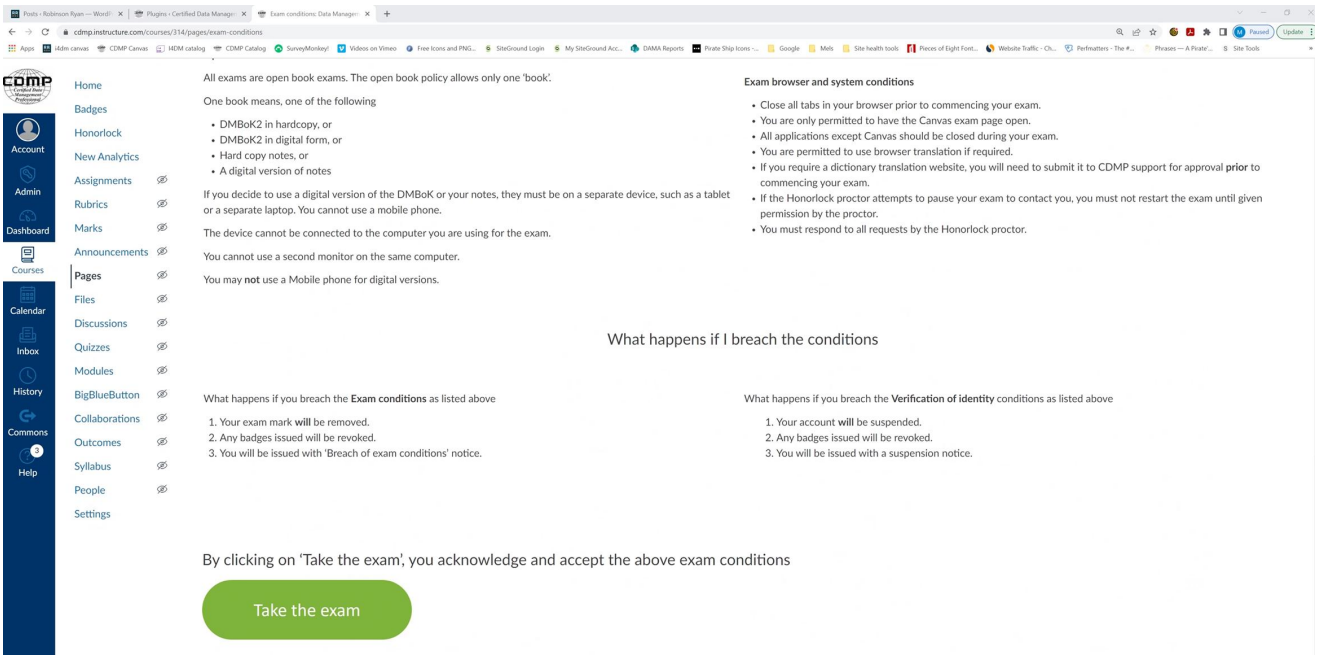




DMA
DATA MANAGEMENT ADVISORS



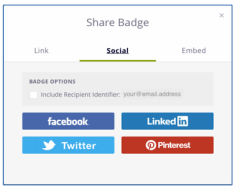
DMA
DATA MANAGEMENT ADVISORS



DMA
DATA MANAGEMENT ADVISORS

CDMP Badges

- Exam platform awards badges for passing CDMP certifications.
- You can put these badges on your CV, LinkedIn, Twitter, Facebook or even Pinterest.
- Badges are not just an image, Badges carry metadata that can be verified.
- EG, you can send your badge to a prospective employer who can verify that your CDMP certification is authentic and valid.
- Badges are part the global movement called the “OpenBadges” specification.
- Started in 2010 by Mozilla, it has expanded to include millions of Open Badges awarded to hundreds of thousands of recipients & supported by certification players such as The Open Group, PMI, Microsoft and IBM.
- Part of the open badges specification & community are public verification authorities that you can use to independently verify your badges are authentic as required.
- In addition to CDMP badges, accumulate badges for certifications including TOGAF, Prince II, MCSE, PMP & PMI.



DMA
DATA MANAGEMENT ADVISORS

DMBoK2 and CDMP® Preparation

Live and On-Demand Classes

Data Management Fundamentals
https://www.dataversity.net/dmbok-and-cdmp-preparation-learning-plan/

Data Governance
https://training.dataversity.net/learning-paths/dgs0-dmbok-and-cdmp-preparation-data-governance-specialist-learning-plan

Data Modelling
https://training.dataversity.net/learning-paths/dms0-dmbok-and-cdmp-preparation-data-modeling-specialist-learning-plan

Data Quality Management
https://training.dataversity.net/learning-paths/dqs0-dmbok-and-cdmp-preparation-data-quality-specialist-learning-plan

Master & Reference Data Management
https://www.dataversity.net/

Metadata Management
https://www.dataversity.net/

- CDMP Online learning program
- Approved by DAMA-I
- Part of DATAVERSITY training center
- Based on DMBoK CDMP syllabus

20% discount code for DAMA UK
"dmadvisors"

DMA DATA MANAGEMENT ADVISORS

Renewal

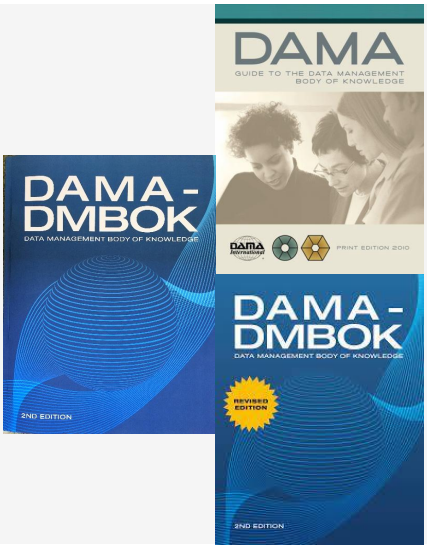
To maintain the validity of your Certified Data Management Professional® (CDMP®) certification, DAMA International® requires all certification holders to participate in an annual renewal process. Initially, the CDMP® certification is valid for three years. To keep your certification active beyond this period, you must:

1. **Annual Maintenance Fee:** **Pay the CDMP® maintenance fee** each year.
2. **Data Management Activity Attestation:** Submit the **Data Management Activity Attestation Form** annually, confirming your ongoing involvement in data management activities.
3. **Code of Ethics:** **Sign and adhere to the current version of the DAMA International® Code of Ethics**, reflecting your commitment to professionalism and ethical standards in the field.

DMA DATA MANAGEMENT ADVISORS

What Is the DAMA-DMBOK Guide?

- The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK Guide)
- A book published by DAMA-I, 406 pages(V1) 624 pages (v2) (also on CD & PDF)
- Available from TechnicsPublications.com or Amazon.com
- Written and edited by DAMA members
- An integrated primer: "definitive introduction"
- Modeled after other BOK documents:
  - PMBOK (Project Management Body of Knowledge)
  - SWEBOK (Software Engineering Body of Knowledge)
  - BABOK (Business Analysis Body of Knowledge)
  - CITBOK (Canadian IT Body of Knowledge)



DMA
DATA MANAGEMENT ADVISORS

DAMA-DMBOK Guide Goals

To develop, build consensus and foster adoption for a generally accepted view of data management.

To provide standard definitions for data management functions, roles, deliverables and other common terminology.

To identify “guiding principles”.

To introduce widely adopted practices, methods and techniques, without references to products and vendors.

To identify common organisational and cultural issues.

To guide readers to additional resources.

DMA
DATA MANAGEMENT ADVISORS

Data Management Disciplines
(DMBoK 2)

- Data Management covers considerably more than Documents and Records management
- Data Governance is at the centre
- Data Management is a **professional** discipline which requires training & awareness

DMA
DATA MANAGEMENT ADVISORS

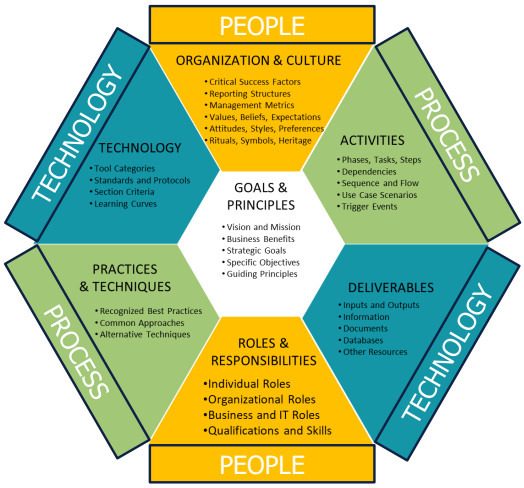
Business Environmental Factors

- Data Management does not exist in a vacuum
- Many environmental factors cross all of the disciplines
- Goals and principles at the centre:
  - Provide guidance how to execute activities and effectively use the tools for successful data management

DMA
DATA MANAGEMENT ADVISORS

Business Environmental Factors

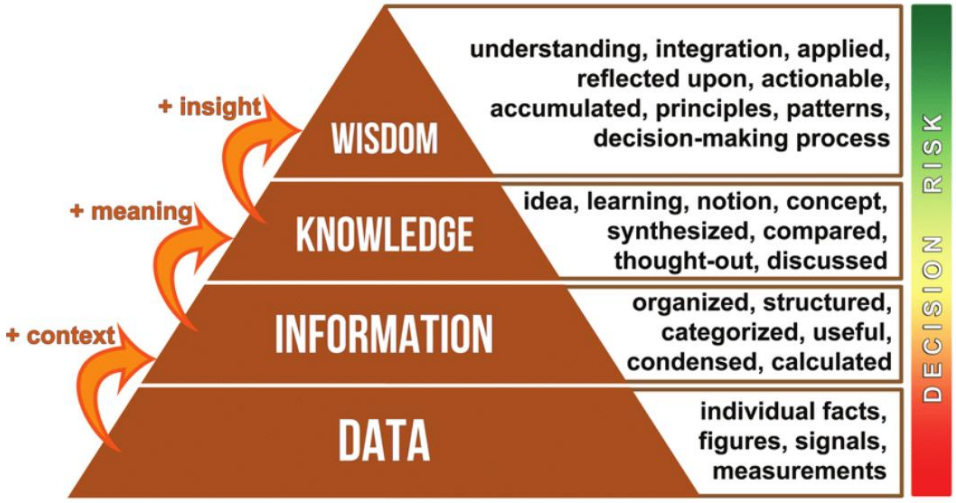
It's NOT just about technology



39

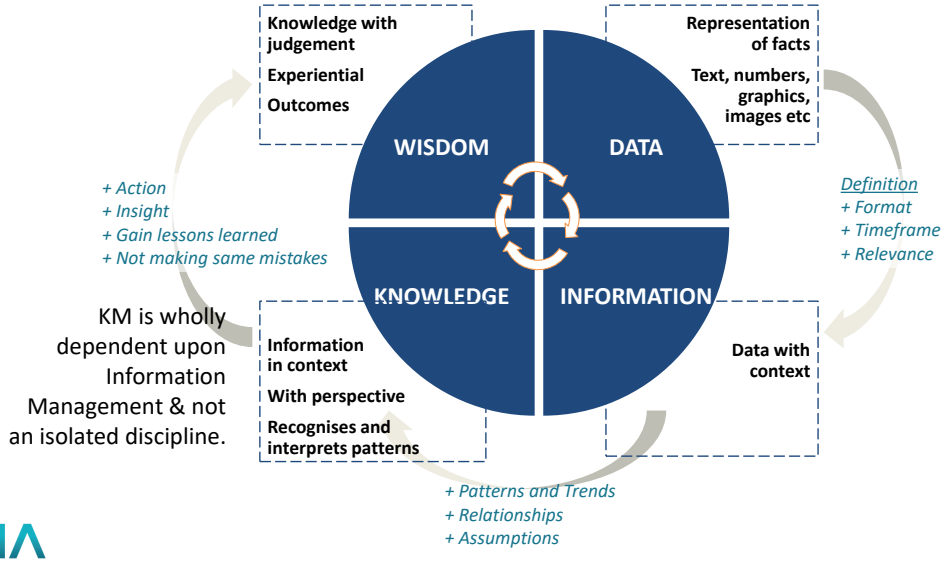
DMA
DATA MANAGEMENT ADVISORS

The Information Value Chain



DMA
DATA MANAGEMENT ADVISORS

Data, Information and Knowledge *(Information Value Chain)*

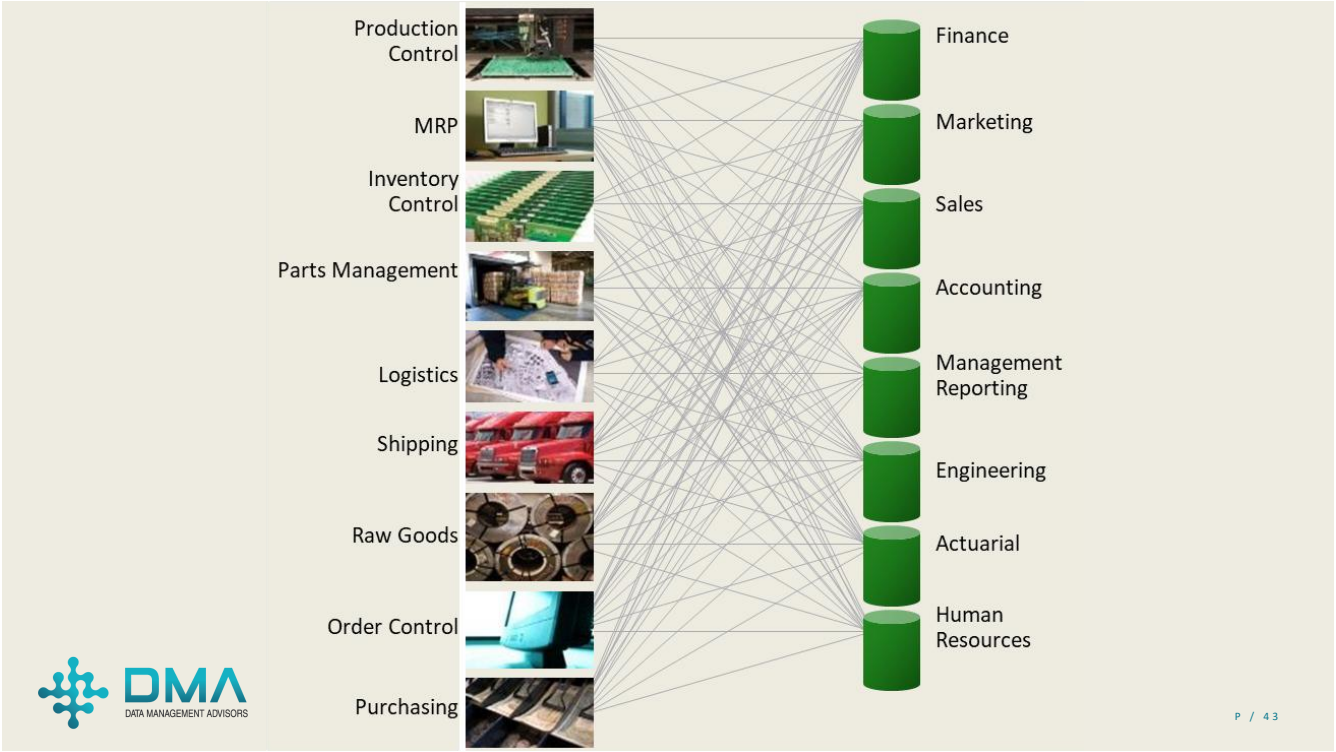


KM is wholly dependent upon Information Management & not an isolated discipline.

DMA
DATA MANAGEMENT ADVISORS

Why is Data Management critical?

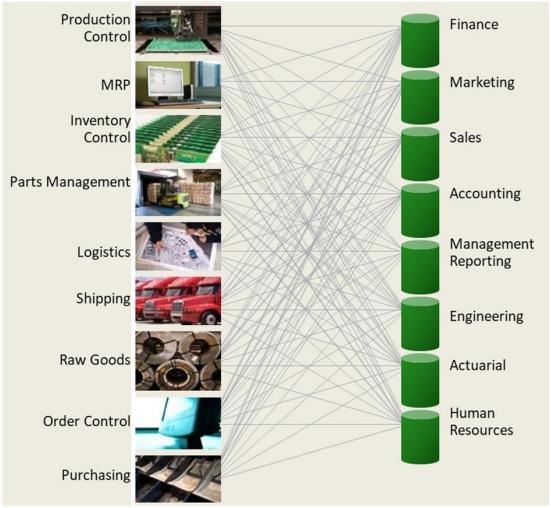
- Higher volumes of data generated by organisations (raw data, devices, CRM, ECM, IOT)
- Proliferation of application-centric systems
- New product development
- To make the management of information front and centre and part of the culture
- Greater demand for reliable information: Gain deep insights through analytics
- Trust in Information: *"What do you mean by ….?"*
- Tighter regulatory compliance
- Competitive advantage: Improved decision making
- Business change is no longer optional – it's inevitable: Agility AND ability to respond to change
- Big Data + AI & ML explosion (and hype)



What is the problem?

Legacy Applications Centric focus = **CHAOS**

The root cause of the muddled state of Information Architecture in most large organisations today is the prevailing **application-centric** mindset that gives applications priority over data, where data is treated as a "by-product"

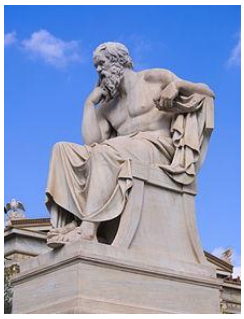


Until we recognize and take action on the core problem, the situation will continue to deteriorate.

DMA DATA MANAGEMENT ADVISORS

Clarity & Definition is vital

"The beginning of wisdom is the definition of terms."



Socrates: 470 – 399 BC

DMA DATA MANAGEMENT ADVISORS

The Problem

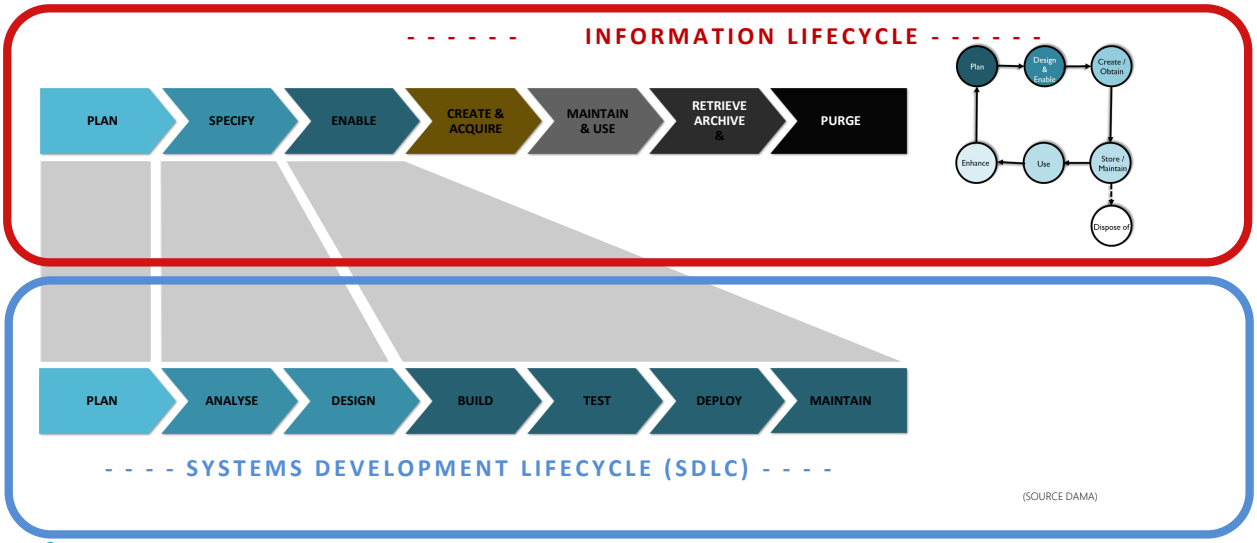
+ vendor promises
(*with business willingness to believe*)
in “next new thing”

There's NO SILVER BULLET

They ALL depend on reliable data … and that’s the bit that’s sadly lacking

DMA
DATA MANAGEMENT ADVISORS

Information Lifecycle & SDLC



DMA
DATA MANAGEMENT ADVISORS

Data is precious

"Data is a precious thing and will last longer than the systems themselves."

Sir Tim Berners-Lee:
Inventor of the World Wide Web.

DMA
DATA MANAGEMENT ADVISORS

Additional / Optional Study Materials

- DAMA-DMBOK (Technics Publishing) (Required)
- Systems Analysis and Design (Kendall and Kendall - Prentice Hall)
- Master Data Management (Loshin - Morgan Kauffman)
- IT Project Management (Schwalbe - Course Technology)

Ref	Question	A	B	C	D	E
GEN2	Information is	Data in context	A management discipline	Always stored in a computer system	A byproduct of IT systems	An abstract concept
GEN3	Information needs to be managed because	It is an asset of the organization	The volumes are large	It contains financial facts	It is stored in Database systems	Processes use it
GEN4	Data differs with regards to other assets because	It uses automation	It can be used yet still retain value	It has value	It is big	It is regulated
GEN5	The Information Lifecycle	Has the same stages as the Systems Delivery Lifecycle	Is used primarily for Data archiving	Is only important in regulated industries	Exists beyond the Systems Delivery Lifecycle	Is not relevant in an Agile environment
GEN6	The DAMA Wheel contains	Knowledge areas	data management processes	data strategy initiatives	maturity model dimensions	data management deliverables
GEN9	Which is a valid DMBoK Environmental component of data management?	Motivation	Hardware Management	Practices & Techniques	Project Management	Database Management.

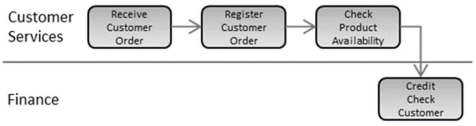
Data Quality Management

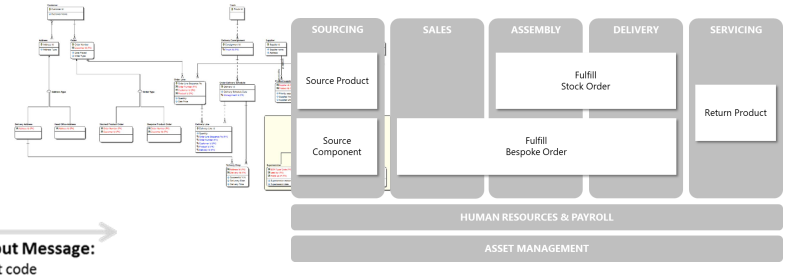
Data Management Process	2%
Big Data	2%
Data Architecture	6%
Document & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

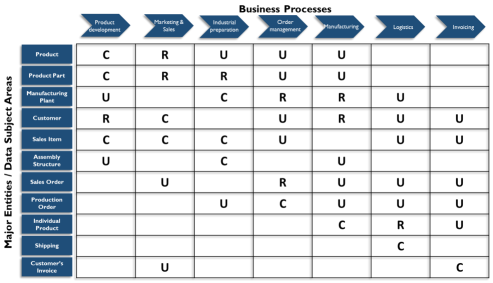
DMA
DATA MANAGEMENT ADVISORS

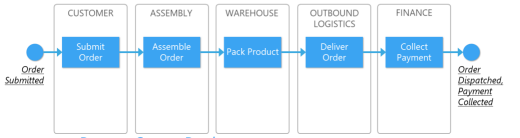
DMA
DATA MANAGEMENT ADVISORS

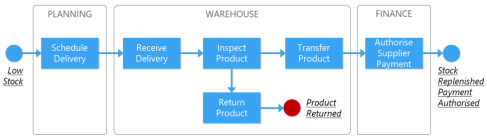
Do we know *everything* that the data might be used for?





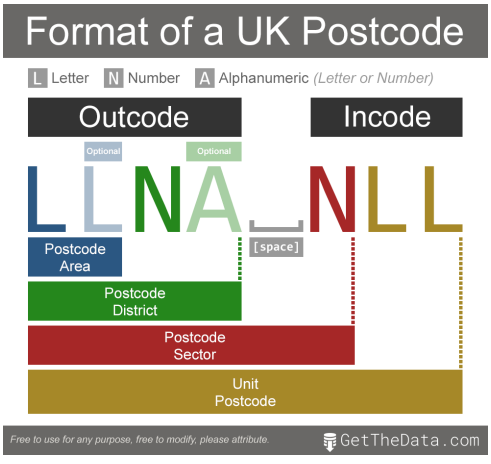


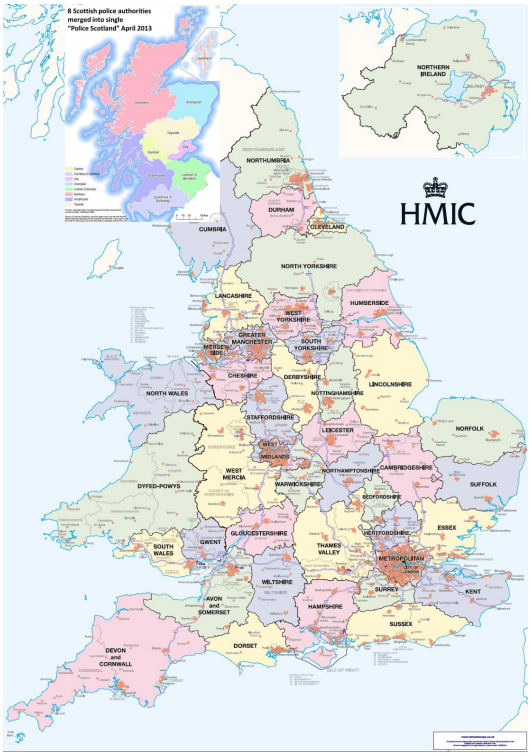




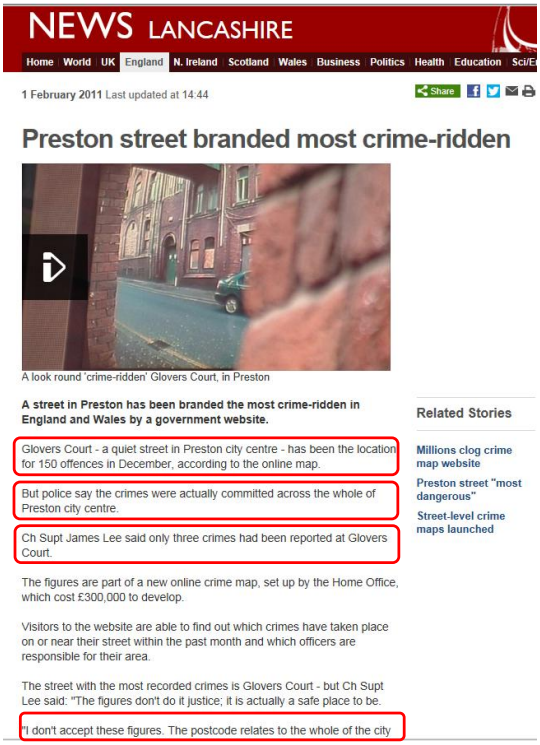
Data Quality Metrics …. Beware, this is NOT an excuse for omitting requirements analysis !

DMA
DATA MANAGEMENT ADVISORS





DMA
DATA MANAGEMENT ADVISORS



NEWS LANCASHIRE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/E

1 February 2011 Last updated at 14:44

Share

Preston street branded most crime-ridden

A look round 'crime-ridden' Glovers Court, in Preston

A street in Preston has been branded the most crime-ridden in England and Wales by a government website.

Glovers Court - a quiet street in Preston city centre - has been the location for 150 offences in December, according to the online map.

But police say the crimes were actually committed across the whole of Preston city centre.

Ch Supt James Lee said only three crimes had been reported at Glovers Court.

The figures are part of a new online crime map, set up by the Home Office, which cost £300,000 to develop.

Visitors to the website are able to find out which crimes have taken place on or near their street within the past month and which officers are responsible for their area.

The street with the most recorded crimes is Glovers Court - but Ch Supt Lee said: "The figures don't do it justice; it is actually a safe place to be.

"I don't accept these figures. The postcode relates to the whole of the city

Related Stories

Millions clog crime map website

Preston street "most dangerous"

Street-level crime maps launched

BBC Sign in | News | Sport | Weather | iPlayer | TV | Ra

NEWS UK

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Env

1 February 2011 Last updated at 16:44

Share

Millions jam street-level crime map website

COMMENTS

A new crime-mapping website for England and Wales is experiencing a "temporary problem" as millions of people log on every hour, the Home Office has said.

Hundreds have contacted the BBC website to report problems accessing the site as officials worked to fix the glitch.

The Home Office said **www.police.uk** was receiving up to five million hits an hour, or some 75,000 a minute.

The maps record reports of different categories of crime and anti-social behaviour

The site allows you to see the offences reported in your local street by entering a street name or postcode.

Home Secretary Theresa May said the maps would give real facts on crime and anti-social behaviour and make police more accountable.

'Complete farce'

In a message on the microblogging website Twitter, the Home Office said: "Hugely popular streetlevel crime maps getting 75,000 hits per minute so you might experience delays. Keep trying."

But some users have reported seeing only a blank page when visiting the website, or a message saying "no police area is associated with this address" when entering a postcode or street.

Others have complained of errors in the actual information shown, with some "quiet streets" next to commercial centres, bars and clubs being tarred with their crimes.

A spokesman for charity Victim Support said it was important that victims of crime had consented as to whether information about their incident was released.

Related Stories

Street branded most crime-ridden

Street-level crime maps launched

Patrolling a city's flashpoints

Seeing the dots relating to 1,672 incidents a mile from my north London front door was quite sobering"

Mark Easton

How Good Does Data Quality Need To Be?

In February 2011, the UK government launched a crime-mapping website for England and Wales (www.police.uk).

Unfortunately, for a number of reasons, the postcode allocated to a specific police incident didn't always correspond to the precise location of the crime.

The net result was that poor accuracy in the recording of geographical information led many quiet residential streets to be incorrectly identified as crime hotspots.

Answer: It depends…

→ *In the context of creating aggregated statistics to assess relative crime rates between counties, the data quality is perfectly acceptable.* ✓ *Data fit for purpose*

→ *However, if the same data is used by an insurance company, there is an issue for the homeowners who receive inflated home insurance premiums.* ✗ *Data not fit for purpose*

Data quality can only be considered within the context of the *intended use of the data*
Data needs to be *"fit for purpose"*
Data quality needs to be assessed on that basis

DMA
DATA MANAGEMENT ADVISORS

Key Points

Asset
- Information is a core business asset
- Its quality can and should be managed like any other asset

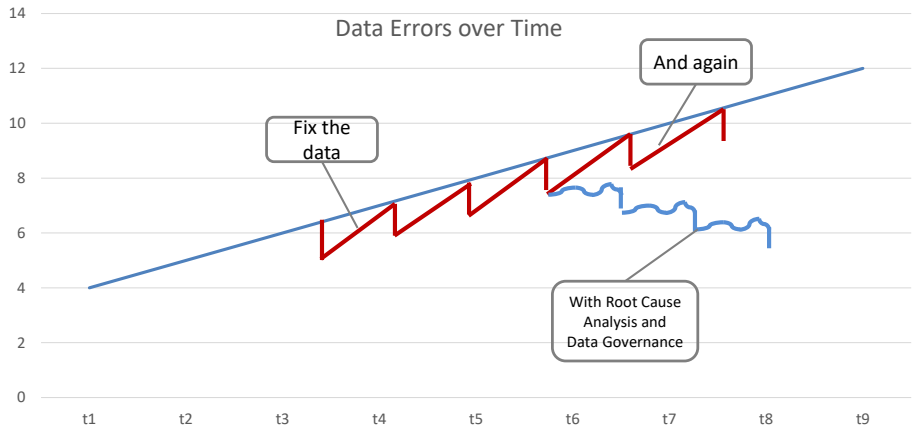
Standards
- Need to define *metadata* like data definition, data type, acceptable values, attribute domains
- Use recognized standards wherever possible

Govern
- Data Quality is managed through appropriate *Data Governance* controls and frameworks
- Quality is in the eye of the down-stream data consumer
- Apply corrections at source, or as close to source as possible

Measure
- Use Statistical Quality control techniques
- Define business rules for conformance to data quality expectations
- Communicate cost of non-quality to help raise awareness

DMA DATA MANAGEMENT ADVISORS

Don't just fix the data



DMA DATA MANAGEMENT ADVISORS

Data Quality Principles (DMBoK 2r)

Criticality: A Data Quality improvement program should focus on the data most critical to the enterprise and its customers. Priorities for improvement should be based on the criticality of the data and on the level of risk if data is not correct.

Standards-driven: All stakeholders in the data lifecycle have Data Quality requirements. To the degree possible, these requirements should be defined in the form of measurable standards and expectations against which the quality of data can be measured.

Objective measurement and transparency: Data Quality levels need to be measured objectively and consistently. Measurements and measurement methodology should be shared with stakeholders since they are the arbiters of quality.

Prevention: The focus of Data Quality improvement program should be on preventing data errors and conditions that reduce the quality of data; it should not be focused on simply correcting records.

Root cause remediation: Improving the quality of data goes beyond correcting errors. Problems with the quality of data should be understood and addressed at their root causes rather than just their symptoms. Because these causes are often related to process or system design, improving Data Quality often requires changes to processes and the systems that support them.

Embedded in business processes: Business process owners are responsible for the quality of data produced through their processes. They must enforce Data Quality standards in their processes.

Systematically enforced: System owners must systematically enforce Data Quality requirements.

Connected to service levels: Data Quality reporting and issues management should be implemented and incorporated into Service Level Agreements (SLA).

DMA
DATA MANAGEMENT ADVISORS

Critical Data

Critical Data Elements are identified by linking them to the Data Quality business drivers; customer experience, effectiveness, and efficiency.

Critical data needs to be managed to a defined Data Quality to meet these objectives.

- Without a particular data element there will be significant impairment to the success of the organization. This impairment size drives a business case for the level of Data Quality Management of a particular data element.
- This value prioritization approach creates a sliding scale of value.

At a defined value level, data elements will become “critical data elements”.

- EG if the data in the customer email address field is incomplete, we will not be able to send product information to our customers via email and we will lose potential sales.
- We know that for every email we send out, we earn $1.00 in revenue.
- This simple example links a clear value driver to Data Quality improvement.

Critical data is often used in:

- Regulatory, financial, or management reporting
- Business operational needs
- Measuring product quality and customer satisfaction
- Business strategy, especially efforts at competitive differentiation.

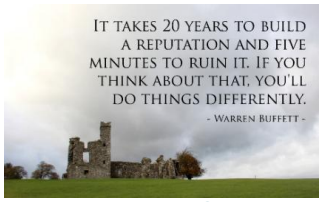
Master and Reference Data is usually critical.

- Data sets or individual data elements can be assessed for criticality based on:
  - the processes that consume them,
  - the nature of the reports they appear in,
  - the financial, regulatory, or reputational risk to the organization if something were to go wrong with the data.

DMA
DATA MANAGEMENT ADVISORS

The Impact of Poor Data

AGGRAVATION



REPUTATION

TELL THEIR FRIENDS

DEATH

RISK

LOSS

https://www.manchestereveningnews.co.uk/news/greater-manchester-news/cracked-gas-main-likely-caused-21353607

https://www.bbc.co.uk/news/av/uk-england-12335044

https://www.theguardian.com/business/2019/may/22/record-44m-fine-for-cadent-after-residents-had-no-gas-for-months

https://www.bbc.co.uk/news/uk-england-lancashire-35048851

http://news.bbc.co.uk/1/hi/scotland/4184962.stm

https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2020/10/ico-fines-british-airways-20m-for-data-breach-affecting-more-than-400-000-customers/

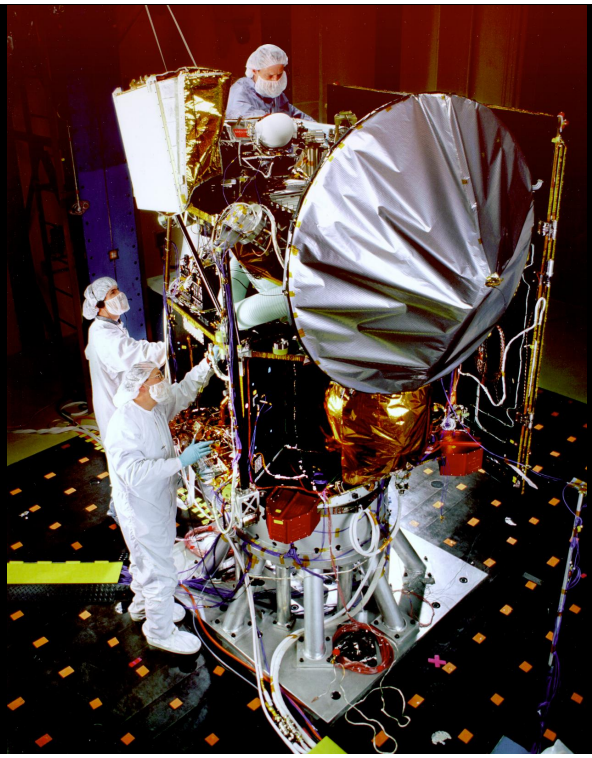
British Airways fined £20m over data breach - BBC News

Google AI chatbot Bard sends shares plummeting after it gives wrong answer | Google | The Guardian

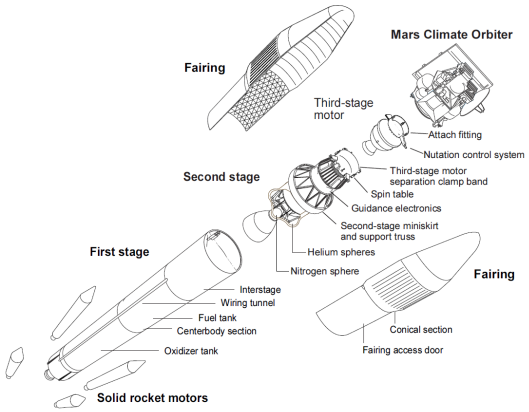
Data being used for the wrong thing: https://www.abc.net.au/news/2023-02-20/robodebt-scheme-government-royal-commission-fraud-income/101998782

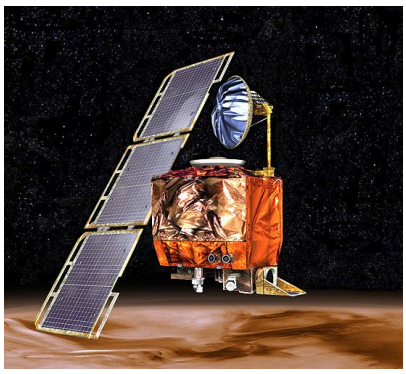
DMA
DATA MANAGEMENT ADVISORS

Launch date: 11 December 1998



Mars Climate Orbiter





Orbiter's Delta II launch vehicle

What Do You Mean By That?

Decay date: 23 September 1999, "Unintentionally deorbited"

Ground software supplied by Lockheed Martin produced results in **pound seconds** ("American"), contrary to its Software Interface Specification (SIS), while a second system, supplied by NASA, used the results expecting them to be in **newton-seconds**, as per the SIS.

The discrepancy between calculated and measured position, resulting in the discrepancy between desired and actual orbit insertion altitude, had been noticed earlier by at least two navigators, *whose concerns were dismissed*.

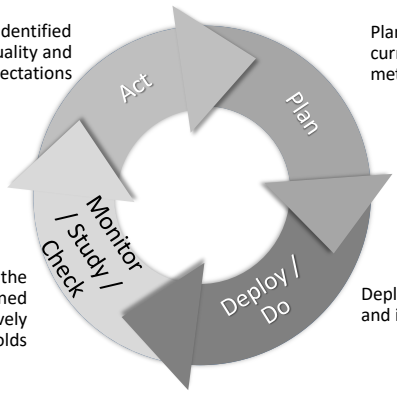
Data Quality Management Cycle

The Data Management Body of Knowledge identifies 4 key activities necessary for operationalising DQM:

Plan, Deploy, Monitor, Act

Acting to resolve any identified issues to improve data quality and better meet business expectations

Planning for the assessment of the current state and identification of key metrics for measuring data quality



Monitoring and measuring the levels in relation to the defined business expectations & actively checking data against thresholds

Deploying processes for measuring and improving the quality of data

Deming built the PDSA method from many problem-solving theories, including Shewhart's specification, production and inspection process

DEMING cycle for continuous improvement
PLAN, DO, STUDY, and ACT

Dr EDWARDS DEMING –
Cycle for Continuous Improvement

The SHEWART cycle:
PLAN, DO, CHECK, and ACT …

WALTER SHEWART –
Cycle for Continuous Improvement

History and comparison of PDSA /PDCA http://www.apiweb.org/circling-back.pdf

The Data Quality Management Approach

Shewart Cycle	Deming Cycle	DQM Cycle	Summary of Tasks
Plan	**Plan**	**Plan**	What data issues are critical to achievement of business objectives? What are requirements for data quality? What are dimensions of DQ and business rules
Do	**Do**	**Deploy**	Conduct data profiling, inspect and monitor for data quality issues; Identify root causes of data quality issues; Apply process or data remediation at, or near, source; Work to eliminate "special causes" of defect
Check	**Study**	**Monitor**	Implement active monitoring of DQ. Apply statistical control. Investigate and understand defects that arise, make plans for remediation
Act	**Act**	**Act**	Take actions to remedy emerging data quality issues as they are identified.

DMA
DATA MANAGEMENT ADVISORS

Juran Trilogy

(Joseph Juran 1908-2008)

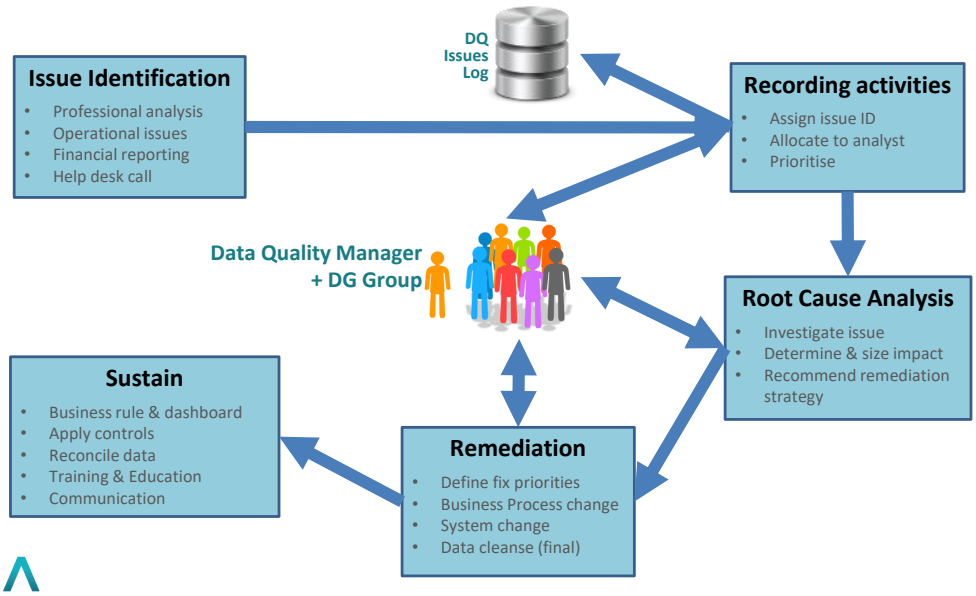
Quality Planning	Quality Control	Quality Improvement
Establish Quality Goals Identify who the customers are Determine the needs of the customers Develop product features that respond to customers needs Develop processes able to produce the product features Establish process control; transfer the plan to the operating forces	Evaluate Actual Performance Compare Actual Performance with Quality goals Act on the difference	Prove the needs Establish the infrastructure Identify the improvement projects Establish the project teams Provide the teams with resources, training and motivation to diagnose the causes and stimulate remedies **Establish controls to hold the gains**

Underlying concept: "managing for quality consists of three universal processes"

- **Quality Planning (Quality by Design)**
- **Quality Control (Process Control & Regulatory)**
- **Quality Improvement (Lean Six Sigma)**

DMA
DATA MANAGEMENT ADVISORS

A simple Data Quality improvement framework



DMA DATA MANAGEMENT ADVISORS

DQ Governance

The steps followed in managing data issues include:

Standardisation — Identifying, capturing, logging, tracking, and updating issues

Assignment — Assignment and tracking of action items

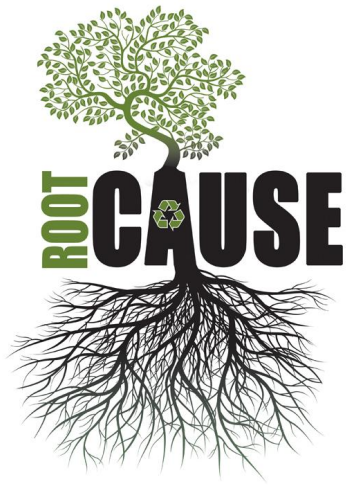
Escalation — Documenting stakeholder viewpoints and resolution alternatives

Completion — Determining, documenting, and communicating issue resolutions

Facilitating objective, neutral discussions where all viewpoints are heard

Escalating issues to higher levels of authority

DMA DATA MANAGEMENT ADVISORS



Step 1: Define the Problem

Step 2: Collect Data

Step 3: Identify Possible Causal Factors

Step 4: Identify the Root Cause(s)

Step 5: Recommend and Implement Solutions

3M

MAN: HUMAN CAUSES – People did something wrong, or did not do something that was needed. Human causes typically lead to physical causes (for example, no one filled the brake fluid, which led to the brakes failing).

METHOD: ORGANIZATIONAL CAUSES – A system, process, or policy that people use to make decisions or do their work is faulty (for example, no one person was responsible for vehicle maintenance, and everyone assumed someone else had filled the brake fluid).

MACHINE: PHYSICAL CAUSES – Tangible, material items failed in some way (for example, a car's brakes stopped working).

DMA
DATA MANAGEMENT ADVISORS

Measuring Data Quality

Dimensions of Data Quality (aka Quality Characteristics) → **What are you going to measure?**

Data Profiling Techniques and Tools → **How are you going to measure?**

Example Dimensions		Example Techniques	
• Completeness	• Timeliness	• Statistical sampling	• Cross-table profiling
• Consistency	• Validity	• Manual profiling	• Statistical Process Control Charts
• Currency	• Accuracy	• Data Profiling tools	• Information flow diagrams (where to measure)
• Precision		• Columnar profiling	• Basic statistics
• Reasonableness		• Intra-table profiling	
• Integrity			

DMA
DATA MANAGEMENT ADVISORS

Data Quality Profiling – Assess Data Quality

Identify a data set for review → Catalogue the business uses of that data set → Subject the data set to empirical analysis using data profiling tools and techniques → List all potential anomalies → Prioritise criticality of important anomalies in preparation for remediation actions and defining data quality metrics

For each anomaly: Review the anomaly with a subject matter expert to determine if it represents a true data flaw → Evaluate potential business impacts

Typical outputs from Data Quality Profiling

- % of the records populated.
- The number of data values populating each data attribute.
- Frequently occurring values.
- Potential outliers.
- Relationships between columns within the same table.
- Relationships across tables

Why Data Quality Profiling?

- Reviewing and refining business policies provides a "top down" view of Data Quality requirements, but a "bottom up" view is crucial to identify existing issues within the data
- This is achieved through an activity known as Data Quality Profiling
- To conduct Data Quality Profiling as efficiently and repeatedly as possible, a specialist Data Quality tool is normally employed
- The result is an invaluable insight into the real operational data, revealing hidden characteristics, patterns and anomalies

Activities:
1. **Define High Quality Data (P)**
2. **Define a Data Quality Strategy (P)**
3. **Define Scope of Initial Assessment (P)**
   1. Identify Critical Data
   2. Identify Existing Rules and Patterns
4. **Perform Initial Data Quality Assessment (P)**
   1. Identify and prioritize issues
   2. Perform root cause analysis of issues
5. **Identify & Prioritize Improvements**
   1. Prioritize Actions based on Business Impact
   2. Develop Preventative and Corrective Actions
   3. Confirm Planned Actions
6. **Develop and Deploy Data Quality Operations (D)**
   1. Develop Data Quality Operational Procedures
   2. Correct Data Quality Defects
   3. Measure and Monitor Data Quality
   4. Report on Data Quality levels and findings

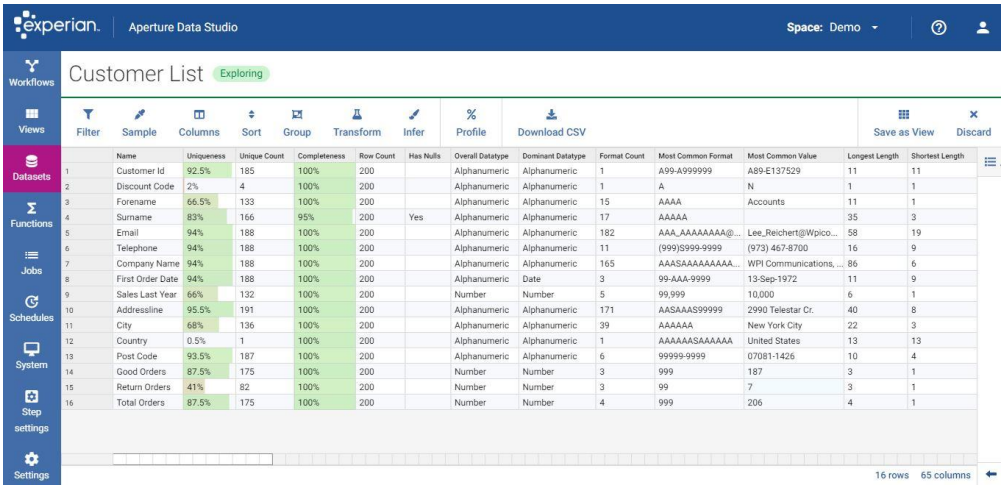
DMA DATA MANAGEMENT ADVISORS

Typical Outputs of Data Quality Profiling

COLUMN PROFILING	• Record count, unique count, null count, blank count, pattern count • Minimum, maximum, mean, mode, median, standard deviation, standard error • Completeness (% of non-null records) • Data type (defined v actual) • Primary key candidates
FREQUENCY ANALYSIS	• Count/percentage each distinct value • Count/percentage each distinct character pattern
PRIMARY/FOREIGN KEY ANALYSIS	• Candidate primary/foreign key relationships • Referential integrity checks between tables
DUPLICATE ANALYSIS	• Identification of potential duplicate records (with variable sensitivity)
BUSINESS RULES CONFORMANCE	• Using a preliminary set of business rules
OUTLIER ANALYSIS	• Identification of possible out of range values or anomalous records

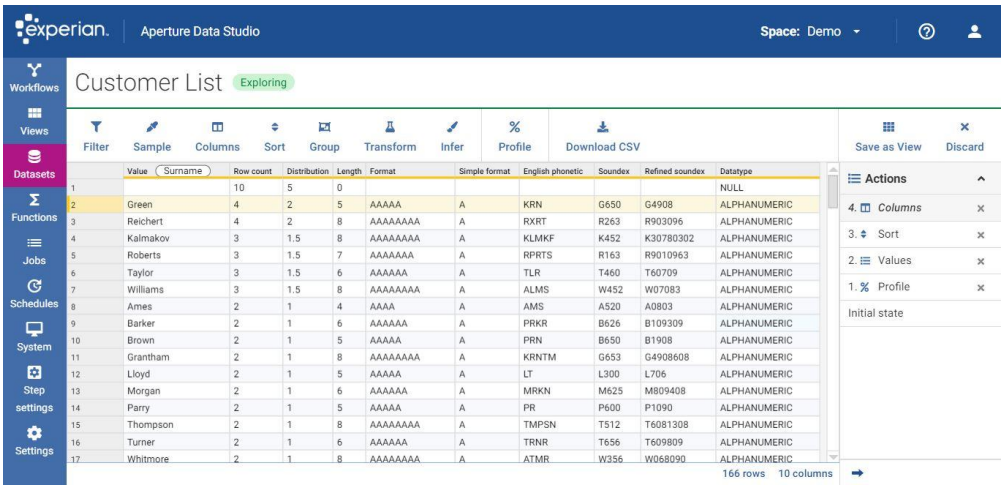
DMA DATA MANAGEMENT ADVISORS

Data Profiling



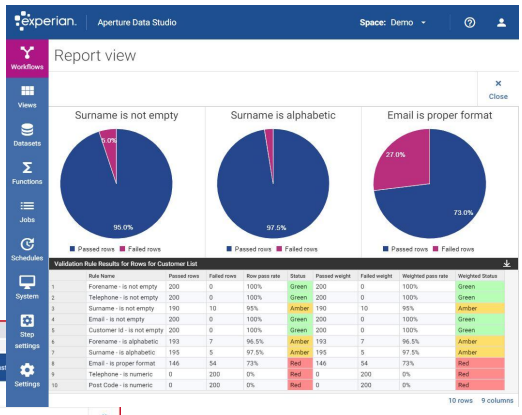
Column Level Stats 1

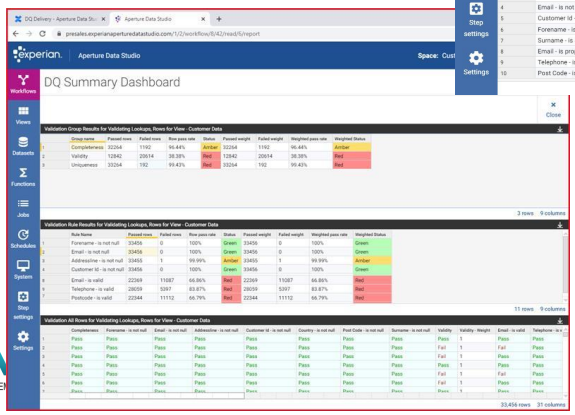
Summary Info



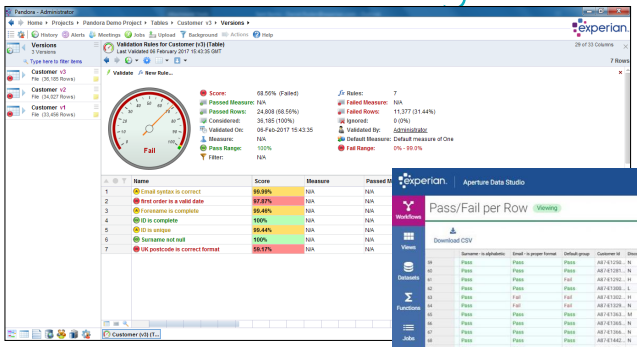
Validation rules

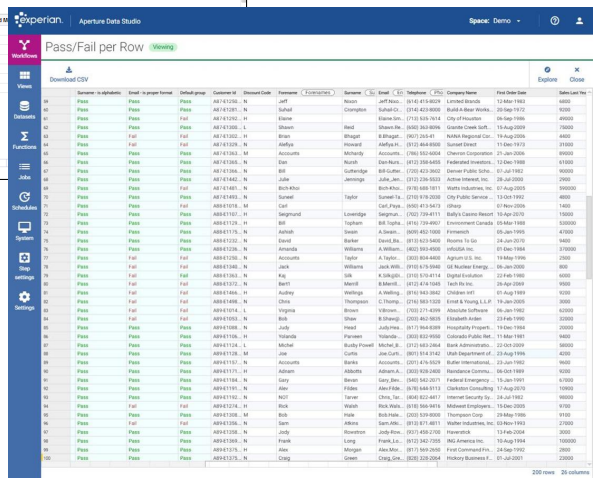
- Every evaluated column can be scored
- Every failing and passing row can be explicitly identified
- Every failure reason is clearly identifiable





Rules-based Monitoring *Validation rules*



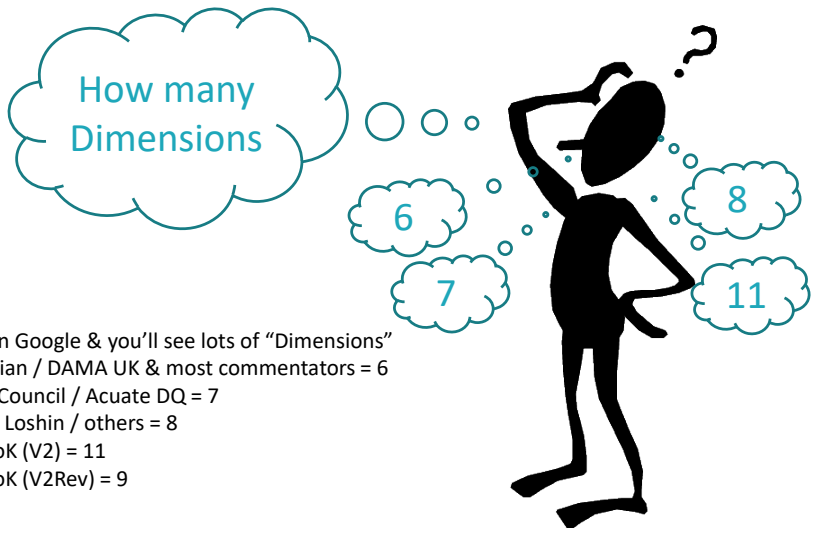


Data Quality Tools and Techniques

Technique	Description	Use
Data Profiling	An objective review of data values against expected standards	Use to measure DQ levels and identify data that is “fit for purpose”
Parsing & Standardisation	Use of patterns and rules to extract data into discrete tokens or to apply standard values based on a defined reference standard	Use to extract data into more granular formats or to identify components in strings; Reduce variation in data set
Transformation	Use of triggered data rules to transform source data to target structures	Use to cleanse/rework data from source to target. Builds on data standardisation rules
Identity resolution & matching	Clustering of records based on “similarity” to identify duplicates or cluster households	Use to dedupe databases or identify important clusters
Enhancement (aka enrichment)	Adds value to data by linking additional information to a record or correcting errors in a record when compared against a definitive source. Can be used to add detail to the record	Use to increase the value of the data, e.g. by appending a market segment to a customer record. Other examples include appending metadata to records to support data lineage tracking or other contextual data.

DMA
DATA MANAGEMENT ADVISORS

Dimensions of Data Quality



Search on Google & you’ll see lots of “Dimensions”

- Experian / DAMA UK & most commentators = 6
- EDM Council / Acuate DQ = 7
- David Loshin / others = 8
- DMBoK (V2) = 11
- DMBoK (V2Rev) = 9

DMA
DATA MANAGEMENT ADVISORS

Data Quality Dimensions DMBoK2 Revised Edition

Quality Attribute	Description
Accuracy	Equal to the "real life" value
Completeness	All mandatory values present, all optional, but implied values present
Consistency	A value in one data set aligns to another data set
Currency	How "fresh" is the data compared to the real-world
~~**Precision**~~	~~The level of detail in the data element~~ covered in Validity & Accuracy
~~**Privacy**~~	~~Does the data need restricted access or monitoring?~~ covered in Metadata classifications
Reasonableness	Is the current behaviour in line with previous behaviour (or average)?
~~**Referential**~~ **Integrity**	All child records must have a parent.
Timeliness	The difference between when the data is available and needed
Uniqueness	Are there duplicates within the data (that do not reflect reality!)
Validity	Is the data within the allowable bounds of the "domain".

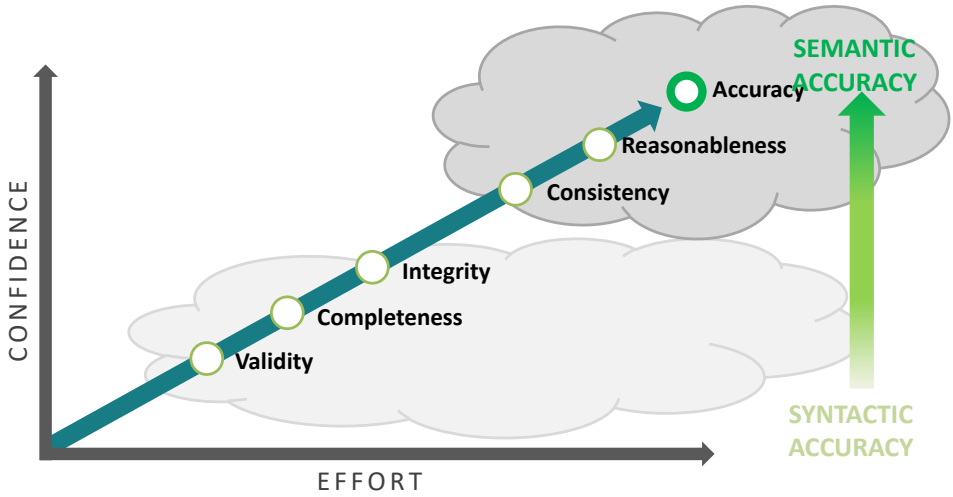
DMA DATA MANAGEMENT ADVISORS

DIM	DESCRIPTION	EXAMPLES
Validity	Validity refers to whether data values are consistent with a defined domain of values.	A domain of values may consist of: • A data type (e.g., text with or without special characters, a number, or a date) • A data range (e.g., a numeric range, date range, or a set of valid text values in a reference table) • A format (e.g., a phone number must contain country, area codes, and the correct number of digits or a currency must have only two decimal places) • The precision of expected values. For example, a date-stamp must be recorded to a millisecond or a number recorded to no more than two decimal places. • The time frame (certain values may also only be valid for a specific length of time, for example, data that is generated from RFID (radio frequency ID) or some scientific data sets).
Completeness	Completeness refers to whether all the required data is present. For example, are all the mandatory components of an address populated, including the street number, street name, city, and country? For example, if the customer account is marked as a debt risk, there must be a debt manager appointed. If the customer account is not marked as a debt risk, there must not be a debt manager appointed.	Completeness can be measured at the column, record, or data set level. • Are columns/data elements populated to the level expected? (Some columns are mandatory. Optional columns are populated only under specific conditions.) • Are records populated correctly? (Records with different statuses may have different expectations for completeness.) • Does the data set contain all the records expected?
Consistency	Consistency is ensuring that data values are coded using the same approach, assessment and valuation criteria. Consistency is between two different data items, potentially within a data set and between data sets, and across time. Consistency may be defined between one set of data element values and another data element set within the same record (record-level consistency), between one set of data element values and another data element set in different records (cross- record consistency), or between one set of data element values and the same data element set within the same record at different points in time (temporal consistency). Consistency can also be used to refer to consistency of format. Take care not to confuse consistency with accuracy or correctness.	Consistency can be measured using the following ways: • Cross record consistency is between data values in the same column – are all the customer addresses recording the head-office details or do some record service delivery centres? • Between associated data sets, where data sets are expected to have values that are linked. For example, when moving data from data capture systems to data warehousing systems, a consistency quality measure will maintain confidence in the values in the data warehouse. • Across time can be applied to any of the previous cases. For example, has the grading of students been the same over time? If a subject is completed previously, would the same learning performance yield the same grade today?

DIM	DESCRIPTION	EXAMPLES
Integrity	Integrity refers to the lack of incoherent values and broken relationships in data.	Integrity can be measured using the following ways: • Coherence – where one data value implies a limited range in another data value, and they are matching. For example, a country has a defined set of states or provinces. If a customer address is in a certain country, they must use a state or province name for the associated set. For example, Alberta is a province of Canada, Pisa is a province of Italy, and Tasmania is a state of Australia. • Parent-child (referential integrity) – where every child must have a parent data value. For example, all the customer address country names must exist on the valid list of countries where the organization is authorized to sell products to. Data sets without integrity are seen as corrupted or have data loss. Data sets without referential integrity have 'orphans' – invalid reference keys, or 'duplicates' – identical rows which may negatively affect aggregation functions.
Timeliness	Data Timeliness refers to the time that data needs to become accessible to a user after its capture or update. Timeliness is the expectation and actual delay before the data becomes available.	For example, An electricity utility needs to maintain its power grid and must have demand data measured and available to a system operator within a few seconds. Otherwise, electricity generation may get out of alignment with demand, causing an energy excess or shortage. As another example, a statistical agency of government produces the Gross Domestic Product (GDP) report two months after the end of the quarter. To meet this timeliness requirement, all the data collections that make up the GDP will have their own specific timeliness requirements.
Currency	Currency is the date that the data was last updated relative to now and the likelihood that it is still correct. Different data sets will have different currency expectations from relatively static data to highly volatile data. Static data remains current for a long period. Volatile data remains current for a relatively short period. For example, country codes are relatively static, remaining current for a long period. A business' bank account is relatively volatile as the balance is constantly changing.	Volatile data, like stock prices on financial web pages, will often be shown with an "as-of-time", so that data consumers understand the risk that the data has changed since it was recorded. During the day, while the markets are open, such data will be updated frequently – it is volatile. Once markets close, the data will remain unchanged but will still be current since the market itself is inactive –it is static. Key terms used in currency include: • Update time – the time stamp of the last update • Volatility measures the rate of change of the date. It may measure across all the data values or measure within a segment. For example, the rate that domestic customers change their address may be different from the rate that corporate customers change their address. • Latency measures the time between when the data was created and when it was made available for use. For example, overnight population of a data warehouse may end up with different data latencies. Data can have a latency of 1 day for data entered into the system early on the prior day, but only a few minutes for data generated just before the load.

DIM	DESCRIPTION	EXAMPLES
Reasonableness	Reasonableness asks whether a data pattern meets expectations. For example, whether a distribution of sales in a geographic area makes sense based on what is known about the customers in that area.	For example, based on previous customer logins at 5pm, are today's customer logins to our systems out of the ordinary? Key terms used in reasonableness include: • Fixed values – this is where the measure is fixed, not dependent on previous behaviour. The maximum currency value of a transaction may be limited in size by the technical limits of a bank's systems and other values bigger than this value is "unreasonable". • Benchmark value – it is possible to measure the average, deviation or any other statistical function and ensure the latest set of values conforms to expected boundaries. For example, there have been between 5,0000 and 10,000 customer address changes per day, but yesterday there were 2,000,000. On a government data collection, the average size company has 20 employees, but the latest returns had an unusual number with an average of 1,500 employees reported.
Uniqueness	Uniqueness states that no real-world entity exists more than once within the data set. Asserting uniqueness of the entities within a data set implies that each row relates to each unique entity in the real world, and only that specific entity.	Uniqueness can be identified using the following techniques: • Key Structure – where there are duplicate keys within the data set. For example, in a data set containing customers, many of which share the same customer number. • Related data – where other data in the data may point to a duplicate. For example, two customers with different customer numbers, but having the same name, date of birth, and office address. • Inquiries or complaints that users cannot access certain data. For example, a customer with multiple services can only access some of their services as the others are recorded under a different customer number (which is a duplicate).
Accuracy	Accuracy refers to the degree that data correctly represents 'real-life' entities. For example, is the person's name in our database actually the person's name in real life? Does the customer actually use that email address? Accuracy is difficult to measure unless an organization can reproduce data collection or manually confirm accuracy of records. Most measures of accuracy use the other dimensions to imply accuracy.	Common techniques to improve accuracy include: • Checking consistency with a data source that has been verified as accurate; a government company register (system of record) or a commercial company Information Broker (system of reference). • Ongoing calibration of devices with reality. For example, over time, weather gauges may drift from providing accurate measurements. An annual program of field inspections visits 10% of the devices and recalibrates them to a standard to ensure they are delivering accurate measurements. • Sampling data values with reality. For example, small number of emails are sent to customers (mainly) to drive sales, but also have the benefit of checking the accuracy of the email address (does it exist, is it used)?

The Path to Accuracy



DMA
DATA MANAGEMENT ADVISORS

ISO 8000 & 22745

ISO 8000 defines quality data as *"portable data that meets stated requirements."* The Data Quality standard is related to the ISO's overall work on data portability and preservation.

Data is considered '*portable*' if it can be separated from a software application. Data that can only be used or read using a specific licensed software application is subject to the terms of the software license.

An organization may not be able to use data it created *unless that data can be detached from the software* that was used to create it.

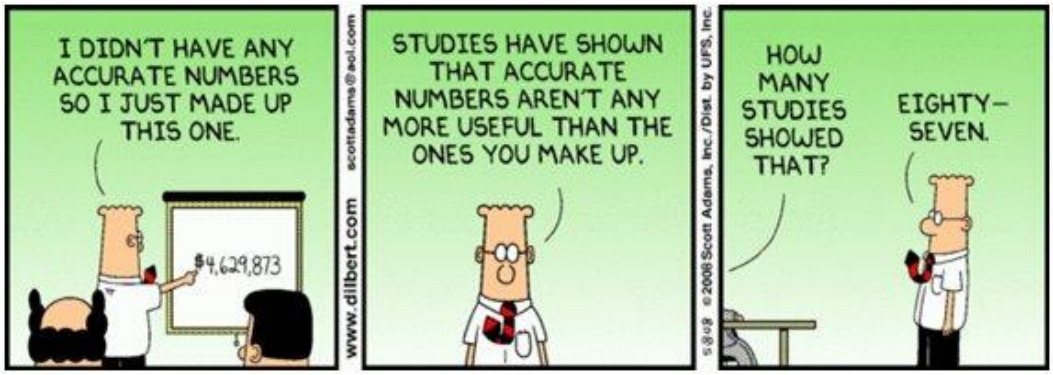
To meet stated requirements requires that these requirements be defined in a clear, unambiguous manner.

ISO 8000 is supported through ISO 22745, a standard for defining and exchanging Master Data.

ISO 22745 defines how data requirement statements should be constructed, provides examples in XML, and defines a format for the exchange of encoded data.

ISO 22745 creates portable data by labelling the data using an ISO 22745 compliant Open Technical Dictionary such as the Electronic Commerce Code Management Association (ECCMA) Open Technical Dictionary (eOTD).

DMA
DATA MANAGEMENT ADVISORS



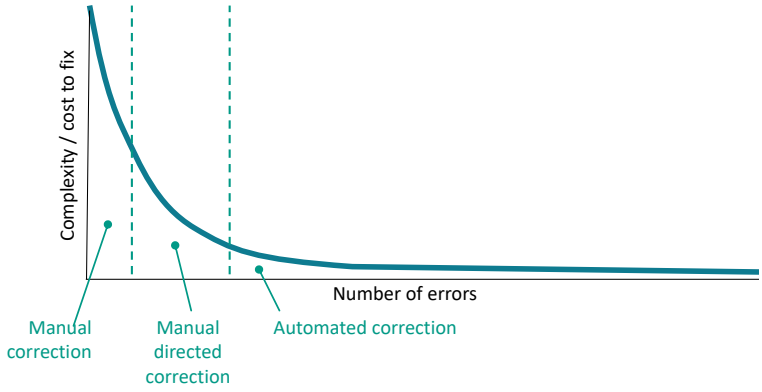
DMA
DATA MANAGEMENT ADVISORS

Data Cleansing Demystified

The quïck fox jump's over the the lazy dog

DMA
DATA MANAGEMENT ADVISORS

Data Correction



Automated correction:

Data transformations and rule-based standardizations, normalizations, reference tables, and corrections.
Modified values are committed *without* manual intervention.

Manual directed correction:

Automated tools cleanse and correct data, up to a point.
Requires *manual* review before committing the corrections to persistent storage.
E.g. scoring mechanism is used to propose a level of confidence in the correction & corrections with scores above a particular level of confidence may be committed without review

Manual correction:

SMEs & Data stewards inspect invalid records and determine the correct values, make the corrections, and commit the updated records.
E.g. Pharmaceutical compound & Clinical trial data

DMA DATA MANAGEMENT ADVISORS

Measures at 3 levels of granularity

Granularity	Common Dimensions	Treatment
Data Element	Validity, Integrity, Consistency, Reasonableness	Edit checks in application Data element validation services Specially programmed applications
Data Record	Consistency, Completeness, Currency	Edit checks in application Data record validation services Specially programmed applications
Data set	Completeness, Uniqueness, Reasonableness	Inspection inserted between processing stages

DMA DATA MANAGEMENT ADVISORS

Caution When Defining Data Quality Indicators

DEFINE DATA QUALITY INDICATORS (DQI) WITH THESE CHARACTERISTICS:

Trackability – make sure each DQI is monitored over time to track progress (NB DMBOK uses the term TRENDING)

Acceptability – make sure it's possible to define what "good" looks like for each DQI – baseline to "goal" = 100%

Relevance – make sure each DQI measures something of importance to the business

Measurability – make sure each DQI can be measured and quantified

Accountability/Stewardship – make sure each DQI links to the Data Governance structure

Controllability – make sure the remedial actions for each DQI are defined

T A R M A C

DMA DATA MANAGEMENT ADVISORS

Common DQ Mistakes

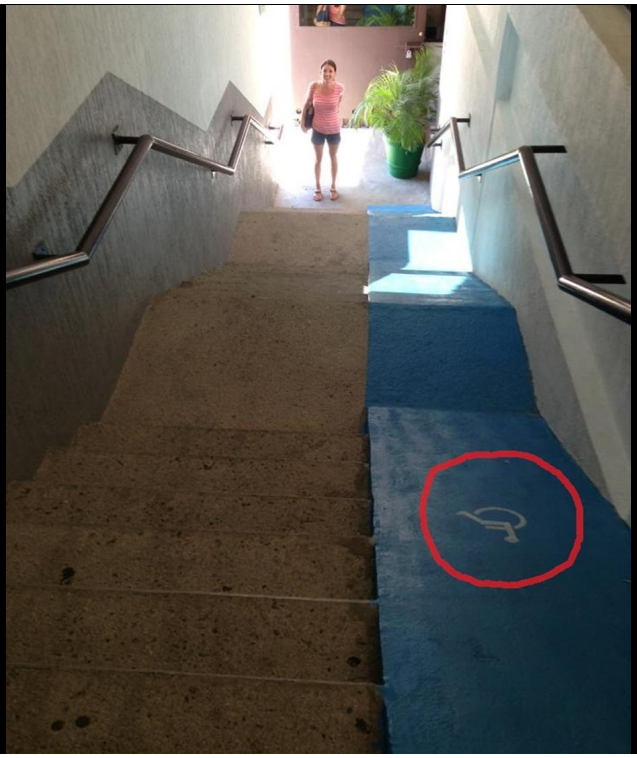
1. **Failing to consider the intended use of the data**
   Data has to be fit for purpose, no more, no less
2. **Confusing validity with accuracy**
   Validity is only the first step towards accuracy
3. **Treating data quality management as a one-time activity**
   Quality data can only be ensured through a continuous cycle
4. **Fixing data in situ vs. at source**
   Clean data for reporting doesn't solve the operational issues of poor DQ
5. **Applying software quality principles to data quality**
   Data is infinitely more volatile than SW and demands a different approach
6. **Blaming systems for bad data**
   People and processes lie at the heart of most DQ problems
7. **Believing that good quality data is the end goal**
   Deriving genuine value through information exploitation is the ultimate aim
8. **Believing Quantity beats Quality for ALL Big Data cases**
   Have you got your "little data" sorted before delving into "Big Data" Quantity doesn't always trump Quality, go in with your eyes open

DMA DATA MANAGEMENT ADVISORS

DMA
DATA MANAGEMENT ADVISORS

...and finally .. It's not always obvious

DMA
DATA MANAGEMENT ADVISORS



Ref	Question	A	B	C	D	E
DQ1	When defining data quality indicators, care must be taken to ensure that they have	Measurability, Relevance & Acceptability	A direct link to the Data Governance strategy	Items in a dashboard showing their improvement over time	The core dimensions of Data Quality	Timeliness, Validity & Accuracy
DQ2	Which of these statements is true?	Data Quality Management is a synonym for Data Governance	Data Quality Management is a continuous process	Data Quality Management only addresses structured data	Data Quality Management is the application of technology to data problems	Data Quality Management is usually a one-off project
DQ3	The Data Quality Management cycle has four stages. Three are Plan, Monitor & Act. What is the fourth stage?	Improve	Prepare	Reiterate	Deploy	Manage
DQ4	Which of these is NOT a typical activity in Data Quality Management?	Defining business requirements & business rules	Analysing data quality	Creating inspection & monitoring processes	Identifying data problems & issues	Enterprise Data Modelling
DQ5	Which of these is NOT an expected role of a Data Quality Oversight Board?	Setting data quality improvement priorities	Establishing communications & feedback mechanisms	Data profiling & analysis	Producing certification & compliance policies	Approving data quality strategies
DQ6	According to DMBoK, which of these is NOT a valid dimension of Data Quality?	Relevance	Timeliness	Currency	Completeness	Reasonableness
DQ7	Which of these is a key process in defining data quality business rules?	De-duplicating data records	Producing data management policies	Separating data that does not meet business needs from data that does	Matching data from different data sources	Producing data quality reports & dashboards
DQ8	A Data Quality Service Level Agreement (SLA) would normally include which of these?	Respective roles & responsibilities for data quality	A Business Case for data improvement	An enterprise data model	Detailed technical specifications for data transfer	A breakdown of the costs of data quality improvement
DQ9	'Top down' and 'bottom up' data analysis and profiling is best done together because	It balances business relevance and the actual state of the data.	It gets everyone involved.	It gives something for the architects to do while the profilers get on with the work	It allows the profiler to show the business the true state of the data	Data quality tools are more productive when they are effectively configured.

AFTER QUIZ 2

1. General (6)
2. Data Quality (9)

Maximum possible score = 15

60% (CDMP Associate) = 9

70% (CDMP Practitioner) = 11

80% (CDMP Master) = 12

Links to Additional Content

Additional Reading:

- *Out of the Crisis*, W. Edwards Deming
- *Data Quality, The Field Guide*, Dr. Thomas C Redman
- *Data Quality, The Accuracy Dimen*sion, Jack Olsen
- *Improving Data Warehouse and Business Information Quality*, Larry P. English
- *Memory Jogger 2,* Brassard et al. *(contains useful summaries of quality management statistical tools)*

Data Storage & Operations

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA DATA MANAGEMENT ADVISORS

Data Storage and Operations (DMBoK 2 revised)

Definition: The design, implementation, and support of stored data to maximize its value.

Goals:
1. Manage availability of data throughout the data lifecycle.
2. Ensure the integrity of data assets.
3. Manage performance of data transactions.

Business Drivers

Inputs:
- Data Architecture
- Data Requirements
- Data Models
- Service Level Agreements

Activities:
1. **Manage Database Technology**
   1. Understand Database Technology (P)
   2. Evaluate Database Technology (D)
   3. Manage and Monitor Database Technology (O)
2. **Manage Database Operations**
   1. Understand Requirements (P)
   2. Plan for Business Continuity (P)
   3. Develop Database Instances (D)
   4. Manage Database Performance (C,O)
   5. Manage Test Datasets (O)
   6. Manage Data Migration (O)

Deliverables:
- Database Technology Evaluation Criteria
- Database Environments
- Migrated / Replicated / Versioned Data
- Disaster Recovery Plan for Data in the Business Continuity Plans
- Database Operations Reports

Suppliers:
- Data Architect
- Data Modeler
- Software Developer
- Application Testing Team

Participants:
- Database Administrator
- Database Technology Specialist
- Data Modeler
- Data Architect

Consumers:
- Data Modeler
- Software Developer
- Application Testing Team
- Infrastructure Operations Team
- Executives and Managers

Technical Drivers

Techniques:
- Test in Lower Environments
- Physical Naming Standards
- Script Usage for All Changes

Tools:
- Data Modeling Tools
- Database Monitoring Tools
- Database Management Tools
- Developer Support Tools

Metrics:
- Data Storage Metrics
- Performance Metrics
- Operations Metrics
- Service Metrics

DMA DATA MANAGEMENT ADVISORS

Key Points

Goals of Data Operations are based on:

- Protecting and ensuring the integrity of data assets
- Ensuring the availability of data throughout its lifecycle
- Optimize performance of database applications

Primary Deliverables

- Production database environments
- Controlled mechanisms and processes for implementation & changes to databases
- Mechanisms for ensuring availability, integrity and recoverability of data
- Mechanisms for detecting and reporting errors
- Mechanisms for ensuring performance according to service level agreements

Main Activities

Database Support

- Primarily undertaken by Production DBAs
- Ensuring the performance and reliability of the database, including performance tuning, monitoring, and error reporting.
- Implementing appropriate backup and recovery mechanisms.
- Implementing mechanisms for clustering and failover of the database (e.g. for 24/7 data availability).
- Implement mechanisms for data archiving.

Data
Technolo
Managem

	2-5 "9s" Availability			
	99%	99.9%	99.99%	99.999%
Daily	14m 24s	1m 26s	8s	1s
Weekly	1h 40m 48s	10m 4s	1m 0s	6s
Monthly	7h 18m 17s	43m 49s	4m 22s	26s
Quarterly	21h 54m 52s	2h 11m 29s	13m 8s	1m 18s
Yearly	3d 15h 39m 29s	8h 45m 56s	52m 35s	5m 15s

DMA
DATA MANAGEMENT ADVISORS

	Hot Data		Warm Data				Cold Data
	Amazon ElastiCache	Amazon DynamoDB	Amazon Aurora	Amazon Elasticsearch	Amazon EMR (HDFS)	Amazon S3	Amazon Glacier
Average latency	ms	ms	ms, sec	ms,sec	sec,min,hrs	ms,sec,min (~ size)	hrs
Data volume	GB	GB–TBs (no limit)	GB–TB (64 TB Max)	GB–TB	GB–PB (~nodes)	MB–PB (no limit)	GB–PB (no limit)
Item size	B-KB	KB (400 KB max)	KB (64 KB)	KB (1 MB max)	MB-GB	KB-GB (5 TB max)	GB (40 TB max)
Request rate	High - Very High	Very High (no limit)	High	High	Low – Very High	Low – Very High (no limit)	Very Low
Storage cost GB/month	$$	¢¢	¢¢	¢¢	¢	¢	¢/10
Durability	Low - Moderate	Very High	Very High	High	High	Very High	Very High
	Hot Data		Warm Data				Cold Data

Factors affecting Availability vs Performance

DMA
DATA MANAGEMENT ADVISORS

Manageability
The ability to create and maintain an effective environment.

Recoverability
The ability to re-establish service after interruption, and correct errors caused by unforeseen events or component failures.

Reliability
The ability to deliver service at specified levels for a stated period.

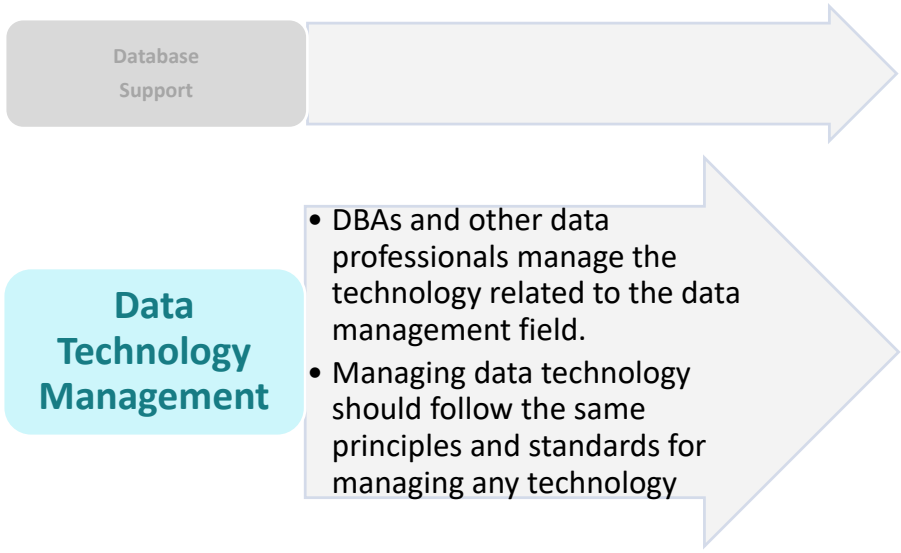
Serviceability
The ability to determine the existence of problems, diagnose their cause and repair/solve the problems.

Causes of poor database performance

DMA
DATA MANAGEMENT ADVISORS

- Memory allocation (buffer/cache for data)
- Locking and blocking
- Failure to update database statistics (SysTables)
- Poor SQL coding
- Poor design of Views (yielding the query from hell)
- Inappropriate indexing
- Application activity
- Increase in the number, size or use of databases
- Database volatility

Main Activities



DMA
DATA MANAGEMENT ADVISORS

Technology Architecture Components - Considerations

Current
Products currently supported and used.

Deployment Period
Products to be deployed for use in the next 1-2 years.

Strategic Period
Products expected to be available for use in the next 2+ years.

Retirement
Products the organisation has retired or intends to retire this year.

Preferred
Products preferred for use by most applications.

Containment
Products limited to use by certain applications.

Emerging
Products being researched and piloted for possible future deployment.

DMA
DATA MANAGEMENT ADVISORS

Ref	Question	A	B	C	D	E
DSO 1	Which of the following activities are performed by data operations staff?	Implement and control database environments, plan for data retention, keep track of database licenses, monitor and tune database performance	Grant access to tables, rewrite SQL statements	Clean data that is of bad quality	Manage the tape libraries	Tune the file systems
DSO 2	The goals of data operations include which of the following?	Assuring the quality of the structured data assets, taking backups and managing security of the database	Assuring availability of the data throughout its lifecycle, protection and integrity assurance of structured data assets and performance optimization of database transactions	Assuring the performance of the network and storage devices, the quality of the SQL statements and the selection of DBMS platform	Assuring backups are taken, managing the performance of SQL and checking data quality	Providing the right database access rights, solving software bugs and managing database logs
DSO 3	The data operations team assures that the data is recoverable by…	Making sure the disks are checked regularly for write errors.	Guaranteeing the applications take proper exports of the data.	Defining and executing the data recovery plan.	Maintaining a test, development and production environment.	Analysing database error logs.

AFTER QUIZ 3

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)

Maximum possible score = 18

60% (CDMP Associate) = 11

70% (CDMP Practitioner) = 13

80% (CDMP Master) = 15

Reference & Master Data Management

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA
DATA MANAGEMENT ADVISORS

DMA
DATA MANAGEMENT ADVISORS

What is Event / Transaction Data?

EVENT DATA EXAMPLE:

"Bob bought a Cadburys Dairy Milk bar from Morrison's on Monday 3rd Jan at 4pm and paid using cash."

WHO	WHAT	WHERE	WHEN	HOW	QUANTITY	AMOUNT
Bob Smith	Dairy Milk bar	Morrison's, Bath	16:00 Monday 3rd January 2022	Cash	1	£7.60

CUSTOMER CODE	PRODUCT CODE	VENDOR CODE	DATE	PAYMENT METHOD	QUANTITY	AMOUNT
BS005	CONF101	WMBATH	2022-01-03 16:00	CASH	1	£7.60

DMA
DATA MANAGEMENT ADVISORS

About Event Data

AKA Transaction data

Describes an action (a verb)
E.g., "buy"

Will include measurements about the action:

- Quantity bought
- Amount paid

Includes information:

Identifying the nouns that were involved in the event
(the Who / What / Where / When / How and maybe even the Why):

- Bob Smith
- Dairy Milk bar
- Morrisons, Bath
- 16:00 Monday 3rd Jan 2022
- Cash

Does not include information:

Describing the nouns:

- Roberta is female, aged 25 and works for British Airways, drives an Audi, ….
- That Dairy Milk is a special offer, extra-large, 850g jumbo bar, half milk, half dark, the recipe is …
- The address of Morrisons Bath is York Place, London Road, Bath, BA1 6AE, the manager is Gary Smith, ...
- Monday 3rd Jan 2022 is a public holiday
- Other methods of payment include store card, debit card, Apple pay …

DMA
DATA MANAGEMENT ADVISORS

What is…

MASTER & REFERENCE DATA?

- Defines and describes the nouns (things) of the business.
  e.g., Customer, Risk, Branch, Loan, Field, Well, Rig, Product, Store, Therapeutic Area, Adverse Event, etc.
- Data about the "things" that will participate in events.
- Provides *contextual information* about events / transactions.
- This Data is stored in ***many*** systems
  - Application package systems
  - Line of Business Systems
  - Spreadsheets
  - SharePoint Lists, …….

DMA
DATA MANAGEMENT ADVISORS

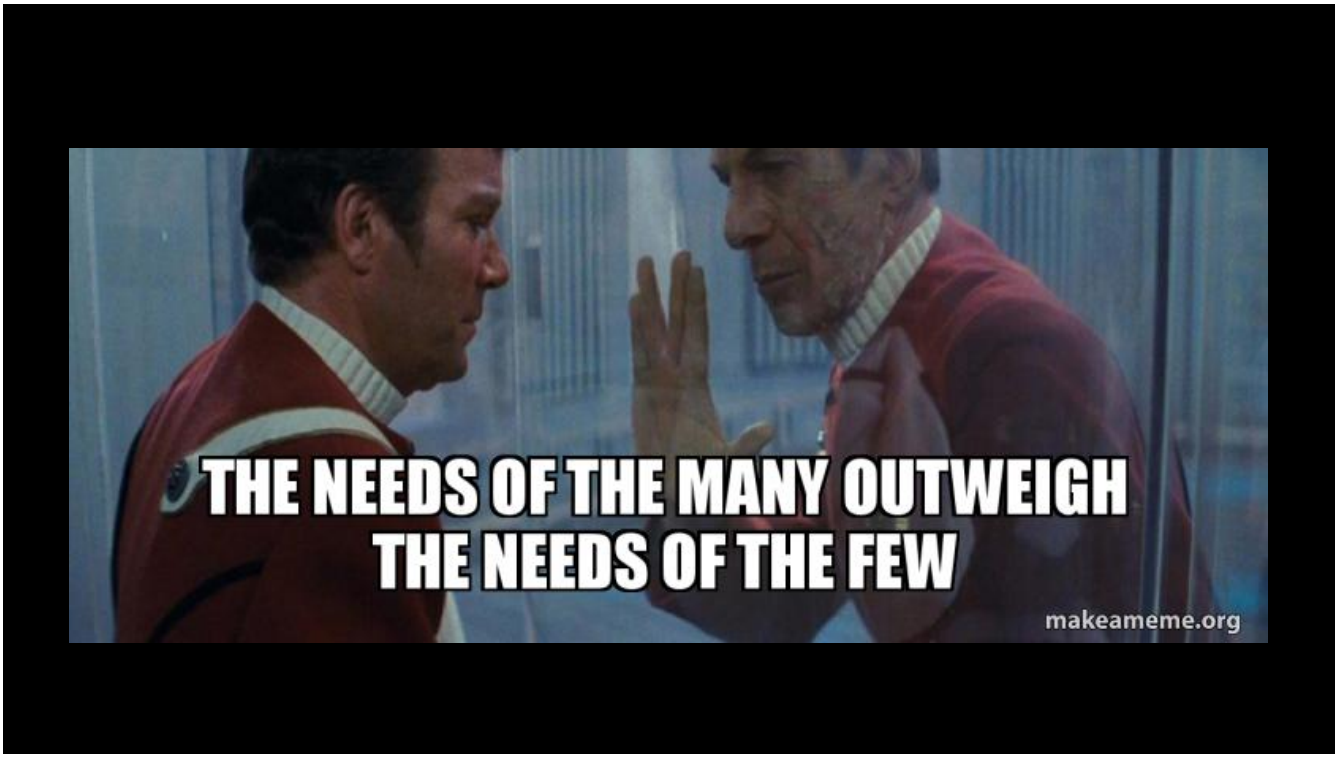
MASTER DATA MANAGEMENT (MDM)?

- The ongoing reconciliation and maintenance of master data.
- Control over master data values to enable consistent, shared, contextual use across systems, of the most accurate, timely, and relevant version of truth about essential business entities.
  (DAMA)
- Comprises a set of processes and tools that consistently defines and manages the non-transactional data entities of an organisation.
  (Wikipedia)

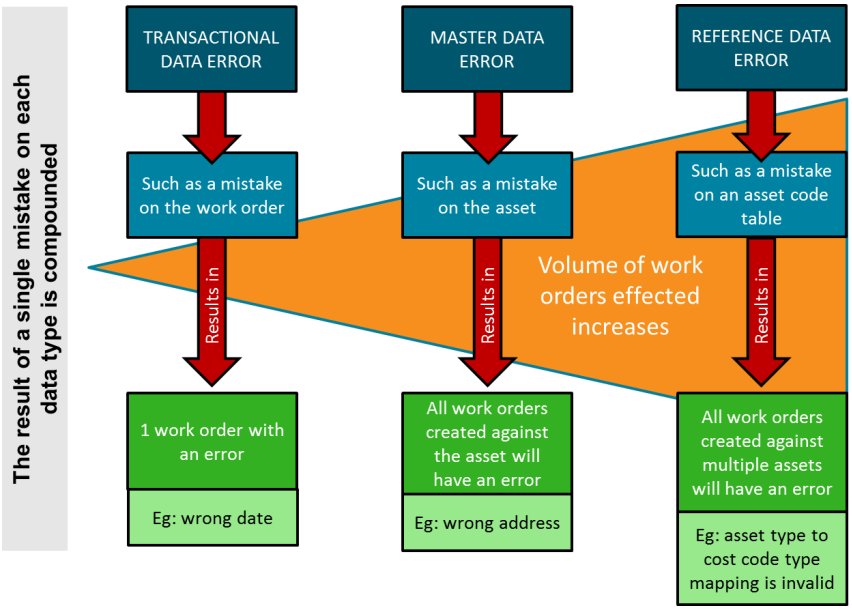
Reference vs Master Data

Characteristic	Reference Data	Master Data
Number of Values	Low, Fixed/Known	Medium-High, Variable/unknown
Volatility	Very low rate of change	Medium-Frequent rate of change
Source	External (mostly); Standards bodies	Internal (mostly)
Number of Attributes per data area	Very low; typically code type/value/description	High
Federation/ownership of attributes per business area	None (all attributes)	Much federation: e.g.. Business Area A masters Attribute 1/Attribute 2, Business Area B masters Attributes 3/4/5
Ease of Governance	Easy	Harder, Higher number of stakeholders
Tool Complexity & Cost	Low (Ref data only) Code (s/w) distribution (e.g. dropdown lists)	High (MDM tools often have ref-data management also)

DMA
DATA MANAGEMENT ADVISORS



Importance of Reference Data



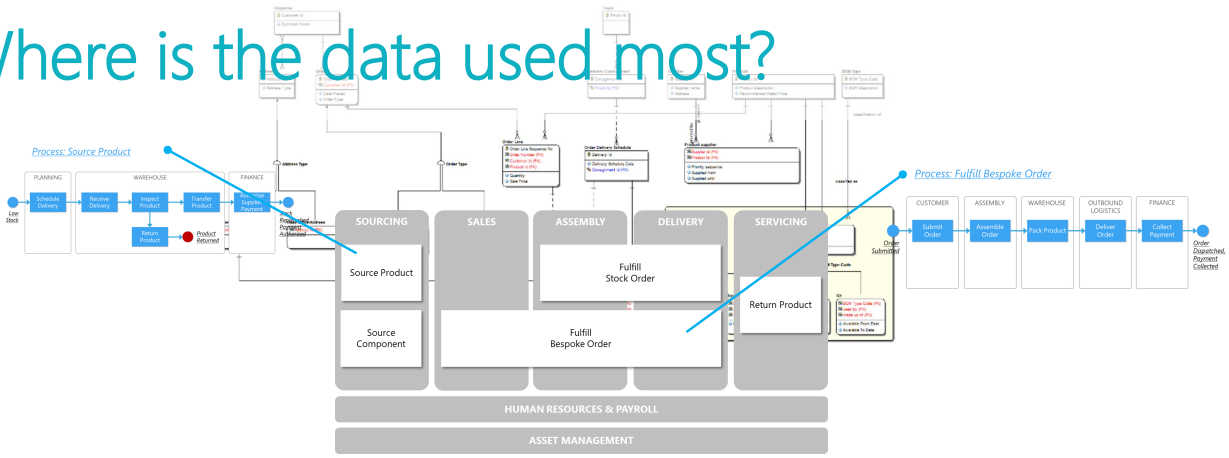
DMA
DATA MANAGEMENT ADVISORS

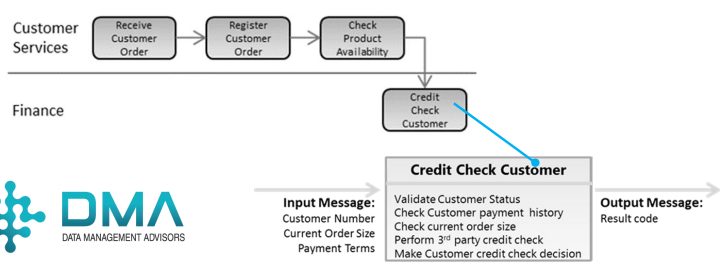
MDM Business Drivers

Meeting Organizational Data Requirements	Managing Data Quality	Managing the Costs of Data Integration	Reducing Risk
Multiple business areas in an organization need access to the same data sets, with confidence that the data sets are complete, current, and consistent.	Data inconsistencies, quality issues, and gaps, lead to incorrect decisions, lost opportunities & inefficiencies;	The cost of integrating new data sources into an already complex environment are higher without Master Data	Master Data can enable simplification of data sharing, interoperability & integration;
"Master" Data often forms the basis of these data sets (e.g., determining whether an analysis includes all ***customers*** depends on having a consistently applied definition of a ***customer***)	Master Data management reduces these risks by enabling a consistent representation of the core, shared entities critical to the organization.	Master Data Management reduces variation in how critical entities are defined and identified.	A Master Data architecture can reduce costs and risk associated with a complex environment

DMA
DATA MANAGEMENT ADVISORS

Where is the data used most?





DMA
DATA MANAGEMENT ADVISORS

Business Processes

Major Entities / Data Subject Areas	Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing
Product	C	R	U	U	U		
Product Part	C	R	R	U	U		
Manufacturing Plant	U		C	R	R	U	
Customer	R	C		U	R	U	U
Sales Item	C	C	C	U		U	U
Assembly Structure	U		C		U		
Sales Order		U		R	U	U	U
Production Order			U	C	U	U	U
Individual Product					C	R	U
Shipping						C	
Customer's Invoice		U				P / 138	C

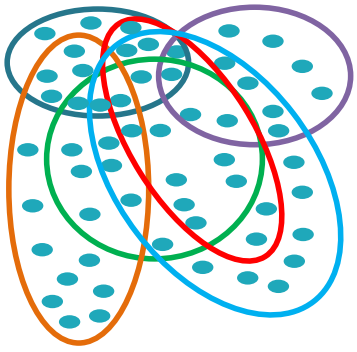
Process and Data ARE related

Create
Read
Update
Delete

Major Entities / Data Subject Areas	Business Processes: Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing
Product	C	R	U	U	U		
Product Part	C	R	R	U	U		
Manufacturing Plant	U		C	R	R	U	
Customer	R	C		U	R	U	U
Sales Item	C	C	C	U		U	U
Assembly Structure	U		C		U		
Sales Order		U		R	U	U	U
Production Order			U	C	U	U	U
Individual Product					C	R	U
Shipping						C	
Customer's Invoice		U					C

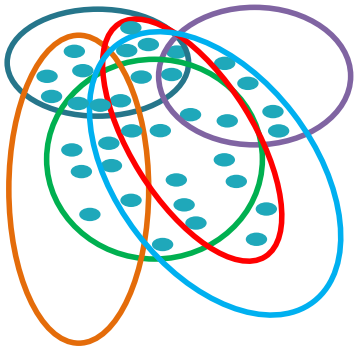
DMA
DATA MANAGEMENT ADVISORS

What should be mastered?



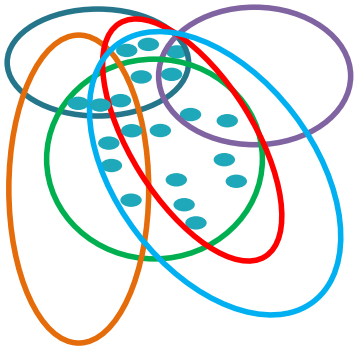
DMA
DATA MANAGEMENT ADVISORS

What should be mastered?



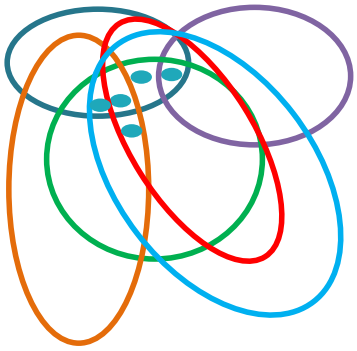
DMA
DATA MANAGEMENT ADVISORS

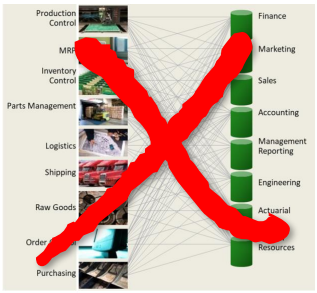
What should be mastered?



DMA
DATA MANAGEMENT ADVISORS

What should be mastered?





DMA
DATA MANAGEMENT ADVISORS

MDM architectures

Standard "Hub" architectures

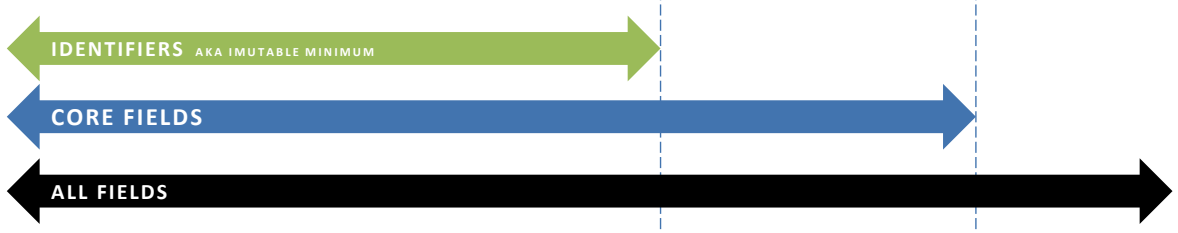
1. **REPOSITORY**
2. **REGISTRY**
3. **HYBRID**
4. **VIRTUALISED**

A key difference is the number of fields that are stored centrally

DMA
DATA MANAGEMENT ADVISORS

Example: Customer

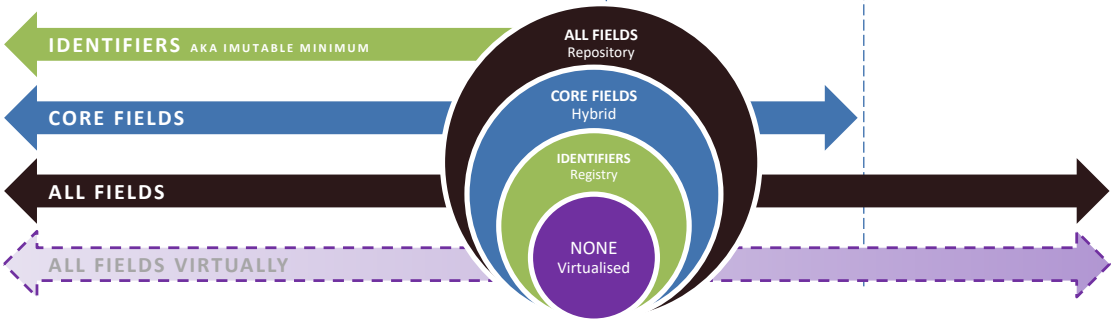
Customer code	First name	Last name	Date of birth	Preferred delivery address post code	Preferred delivery address line 1	Credit rating	Occupation	Car
JBS005	Roberta	Smith	1985-12-25	BA1 7LA	Royal Crescent	A	Information Architect	Audi R8



DMA
DATA MANAGEMENT ADVISORS

Example: Customer

Customer code	First name	Last name	Date of birth	Preferred delivery address post code	Preferred delivery address line 1	Credit rating	Occupation	Car	....
BS005	Roberta	Smith	1985-12-25	BA1 7LA	Royal Crescent	A	Information Architect	Audi R8	



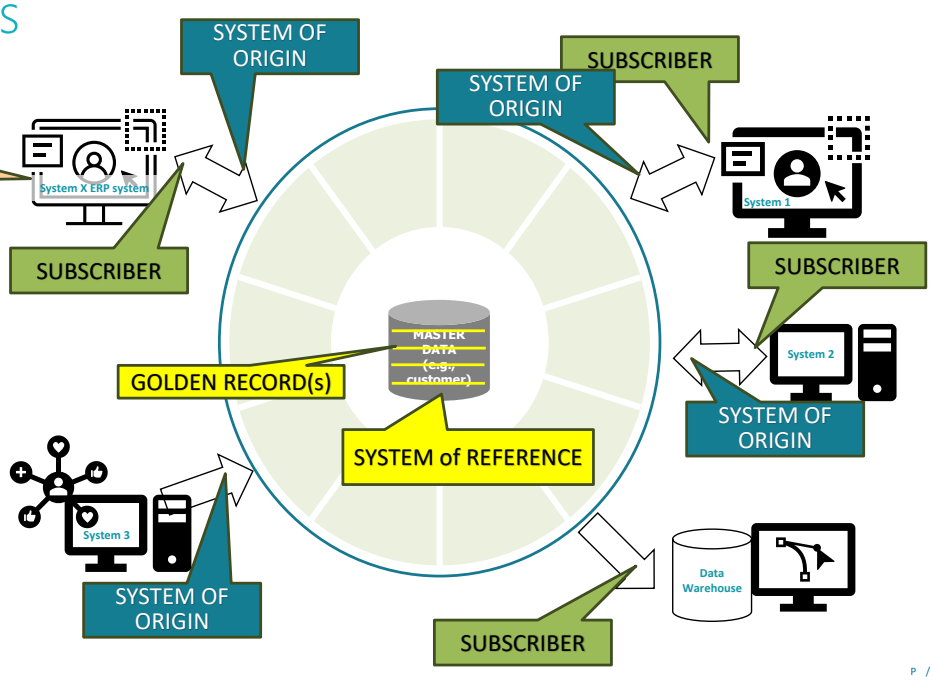
DMA
DATA MANAGEMENT ADVISORS

MDM Terms

SYSTEM of RECORD
(AKA DATABASE of RECORD)

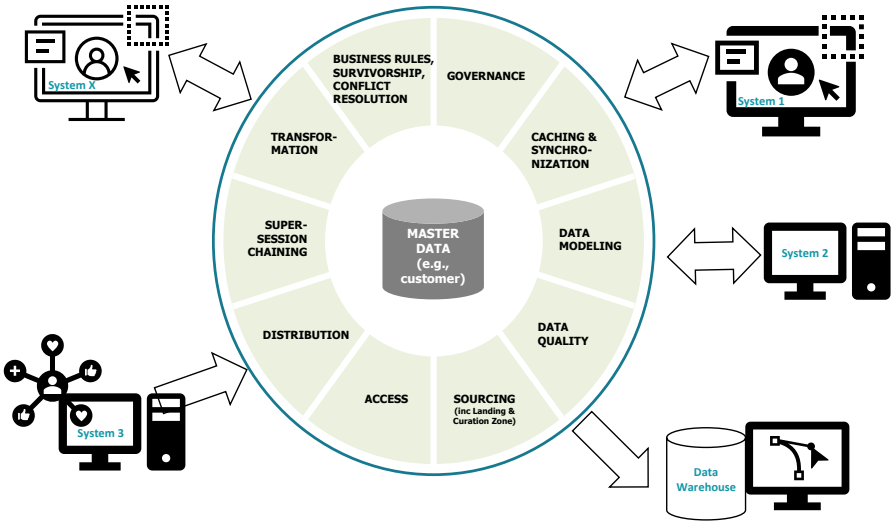
A System of Record is an authoritative system where data is created / captured, and / or maintained through a defined set of rules and expectations (e.g., an ERP system may be the System of Record for sell-to customers).

A System of Reference is an authoritative system where data consumers can obtain reliable data to support transactions and analysis, even if the information did not originate in the system of reference. MDM applications, Data Sharing Hubs, and Data Warehouses often serve as systems of reference.



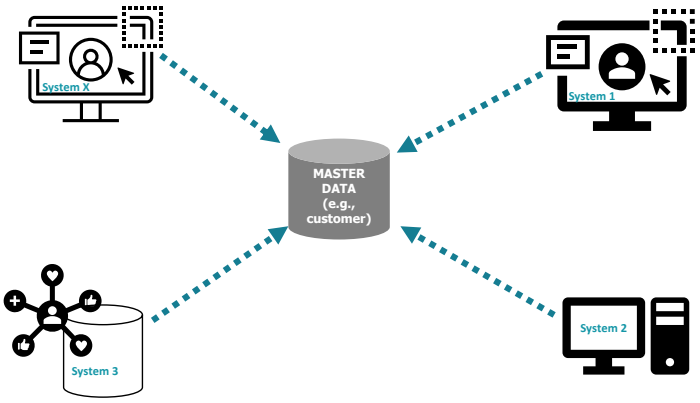
DMA
DATA MANAGEMENT ADVISORS

Typical MDM Components



DMA
DATA MANAGEMENT ADVISORS

MDM Architecture Implementation Styles
REGISTRY

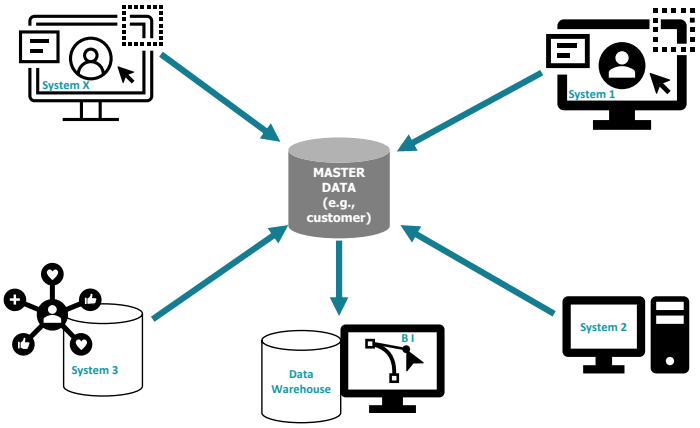


- Low control, autonomous environments
- The most difficult MDM style to implement data governance.
- Non intrusive to edge applications
- Emphasis is on remote data and Application to Application integration
- Distributed governance
- Faster to implement than coexistence & centralised

BENEFITS

- Can be used to analyse the data while avoiding the risk of overwriting information in the source systems.
- Helps avoid potential compliance failure or other regulatory repercussions
- Provides a read-only view of data without modifying master data and is a useful way to remove duplications and gain consistent access to your master data
- Low-cost, rapid data integration with the benefit of minimal intrusion into your application systems.

MDM Architecture Implementation Styles
CONSOLIDATION

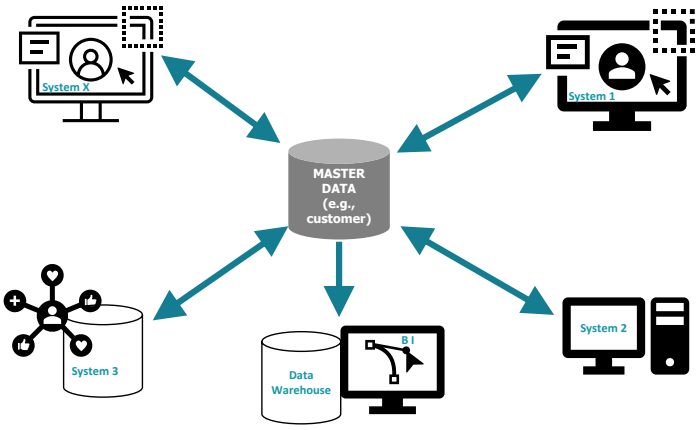


- Ideal for reporting or analytics that reside in a BI / DW environment
- Non intrusive to the business
- BI is the “main” use
- Benefits primarily BI use
- No cleaning of source data
- Higher data latency

BENEFITS

- Pulls master data from several existing systems into a single managed Master Data Management hub.
- Data is cleansed, matched and integrated to offer a complete single record (Golden Record) for one or more master data domains.
- Inexpensive and quick to set up, providing a fast and efficient way to facilitate enterprise-wide reporting.
- Mainly used for analysis, providing a trusted source of data for reporting and analytics.

MDM Architecture Implementation Styles
COEXISTENCE

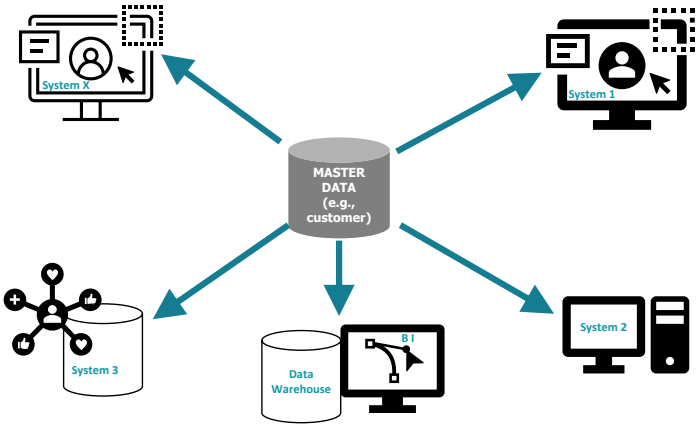


- Large scale distributed model
- Greatest change to information infrastructure
- Greatest need to mirror data
- Global AND local governance
- Risk over control & security
- Focused on shared services
- Lives the “single version of the truth”

BENEFITS

- Data is mastered in source systems and then synchronized with the hub.
- Data can coexist harmoniously and still offer a single version of the truth.
- Quality of master data is improved, and access is faster.
- Reporting is easier as all master data attributes are in a single place.
- Consolidation style can naturally evolve into a Coexistence style if business decides it requires the advantage of linking centrally governed, quality data back to the source systems.

MDM Architecture Implementation Styles
CENTRALISED

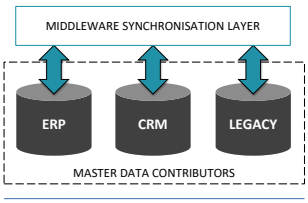


- High control top-down environments.
- Centralised governance - the easiest MDM style to implement data governance based on controls placed on persistent data.
- Greatest change to application infrastructure
- Highly invasive to the business
- Greatest control over access & security
- Focused on common services
- Lives the “single version of the truth”

BENEFITS

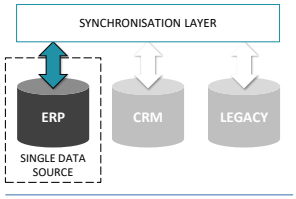
- Master data is accurate and complete at all times.
- Security and visibility policies at a data attribute level can be supported.
- Centralized set of master data for one or more domains.
- MDM now is the system of origin, can leverage powerful data governance capabilities of the MDM.
- Centralized style can evolve from Consolidation or Coexistence styles.

MDM… A Hub is not the only way



Synchronized Master

- Multiple operational systems acting as master data contributors
- Real-time information availability
- Well suited to enterprises where data is stored across multiple source systems
- Well suited to low data velocity operations
- **Like Coexistence**

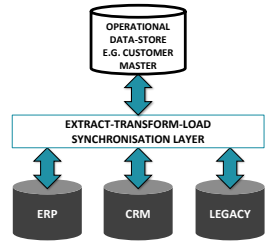


Application Specific Master

- One operational system as master data provider
- Well suited to enterprises where data is primarily stored in a single source system
- Support from many enterprise vendors
- **Hybrid, partially like Coexistence**

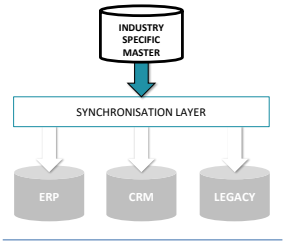
DMA
DATA MANAGEMENT ADVISORS

MDM… A Hub is not the only way



Hub Based Master

- Operational Hub structure that overlaps operational and analytical environments
- Supports concept of an Enterprise Data Warehouse
- Multiple systems acting as data providers
- Appropriate for low data latency & velocity operations
- Requires careful data quality management
- **Coexistence**

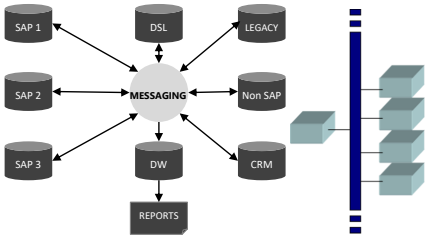


Master Overlay

- Stand-alone, application-neutral master data
- Industry-specific data model
- Well suited to vertical industries in aligning front-office & back-office systems in real time
- Now most often seen for **Reference** Data
- **Centralised**

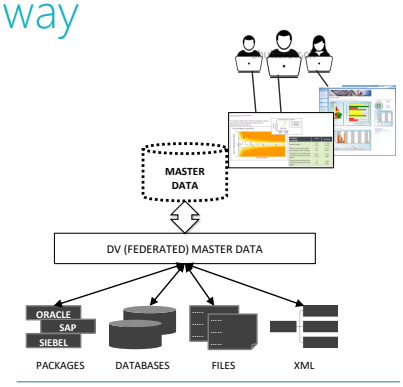
DMA
DATA MANAGEMENT ADVISORS

MDM… A Hub is not the only way



Real Time Data Movement

- Data is moved between the various systems using a messaging integration hub or ESB
- Data movement is done in real time
- Movement of data can be one or two way as required
- **Coexistence via a message-based architecture**



Data Virtualization** (Federated) **Master Data

- Virtual MDM hub created via DV layer
- DV layer informed from Logical Data Model
- Powerful data integration to access multiple disparate systems
- Rapid time to solution
- Outstanding prototyping, proof of value approach
- Compliments other Data Integration approaches
- **Could be Consolidated, Coexistence or Centralised**

Master Data Match Rules

Three primary scenarios:

Rules around the matching, merging and linking of data from multiple systems about the same person, group, place or thing.

1. Duplicate identification match rules

Focus on a specific set of fields that uniquely identify an entity and identify merge opportunities without taking automatic action. Business Data stewards can review these occurrences and decide to act on a case-by-case basis.

Less used

2. Match-merge rules

Match records and merge the data from these records into a single, unified, reconciled and comprehensive record. If the rules apply across data sources, create a single unique and comprehensive record in each database.

Most commonly used

3. Match-link rules

Identify and cross-reference records that appear to relate to a master record without updating the content of the cross-referenced record. Match-link rules are easier to implement and much easier to reverse.

Used extensively in Police / security services, local authorities

.. and Data Enrichment can also help with business entity resolution !

Master Data Matching

Deterministic matching algorithm

- Matches exact character to character of one or more fields **(Exact string match)**
- Has a discrete all or nothing outcome
- All identifiers being matched have equal weight
- Is better suited when there is no great consequence to an error in matching

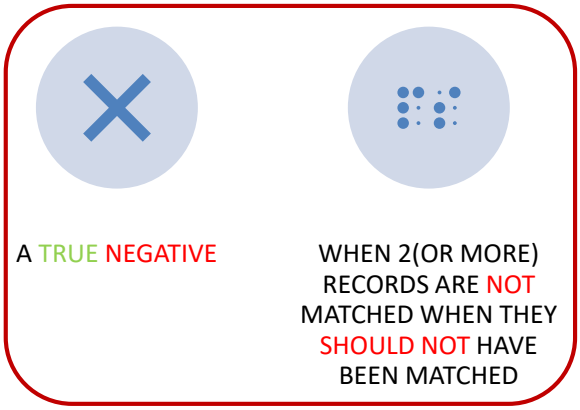
Probabilistic matching algorithm

- Each variable to be matched is assigned a weight based on its discriminating power
- A score is assigned based on weight and degree of match
- Individual attribute matching scores are used to create a match probability percentage
- Following the matching process there are typically records requiring manual review and decisioning

MDM matching algorithms benefit from the data characteristics of:

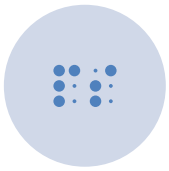
- High validity of the data
- Distinctiveness across the population of data
- High level of comparability of the data elements
- Structural heterogeneity of data elements

Master Data Matching



Master Data Matching

A FALSE NEGATIVE



WHEN 2(OR MORE) RECORDS ARE NOT MATCHED WHEN THEY SHOULD HAVE BEEN MATCHED

A FALSE POSITIVE

WHEN 2(OR MORE) RECORDS ARE CONSIDERED TO BE MATCHED WHEN THEY SHOULD NOT BE

Single Domain & Multi Domain MDM

SINGLE DOMAIN

- E.g. Product (PIM), Customer (UCM), Vendor, Laboratory (LIM)
- Rich Data Model (frequently an Industry data model)
- Incorporate specific “domain” features …
  - Rich functionality & logic for the specific data domain
  - House holding
  - Address chaining
  - Company Hierarchies
  - Tailored data matching (deterministic or probabilistic)
- Engineered to exploit 3rd party data sources
  - GB PAF, USPS Zip+4, Electoral Roll, CCJ …
  - D&B, Liquidations, Companies House, …
  - CAS, COSHH, …
- Interfaces to domain LoB apps

MULTI DOMAIN

- Highly configurable MDM solutions
- Informed by your Data Model
- Fewer specific data domain features
- Standard MDM processes
- Interfaces to mainstream apps
- Results in fewer MDM solutions throughout the Enterprise

Implementation: Operational vs Analytical MDM

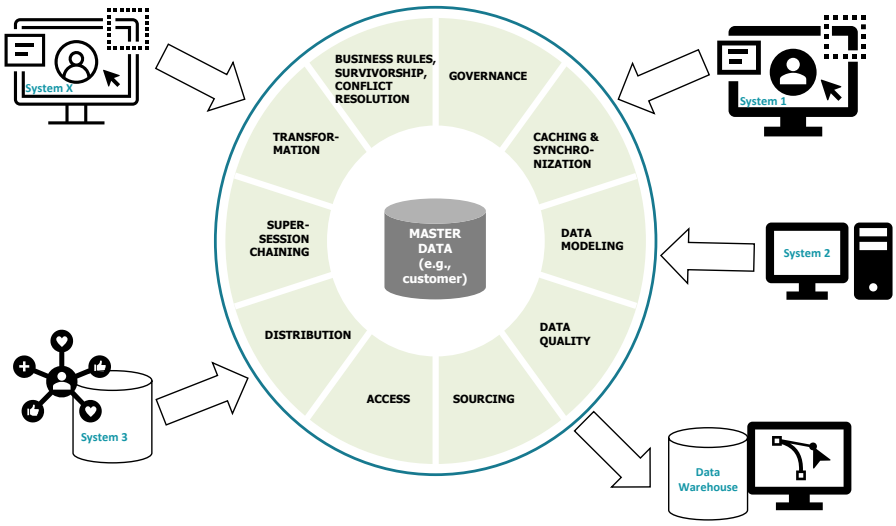
ANALYTICAL

- MDM (usually a hub) created for use in Data Warehouse / Business Intelligence solutions
- Data integration (e.g., ETL) from source systems still required
- Data cleanse, consolidate, merge, de-duplicate, enrich etc. still required
- Master Data used for analytical purposes to ensure consistency of analyses, findings etc.
- Master Data NOT used by live operational line of business solutions, therefore operational systems do not see any benefits of MDM approach
- NOT essential to address business rules (such as survivorship, channels, governance) for operational MDM
- Easier (organizationally) to implement
- Impact of failure only on BI activities

DMA
DATA MANAGEMENT ADVISORS

MDM Implementation

Analytical



DMA
DATA MANAGEMENT ADVISORS

Implementation: Operational vs Analytical MDM

OPERATIONAL

- MDM (whatever is most appropriate architecture) created for use in LIVE operational Business solutions
- Data integration (e.g., ETL) from source systems essential
- Data cleanse, consolidate, merge, de-duplicate, enrich etc. essential
- Master Data is used by the live operational line of business solutions
- Essential to address business rules (such as survivorship, conflict resolution, channels, attribute federation, data governance) for operational MDM
- More difficult (technically & organizationally) to implement
- Impact of failure severe
- Benefits extensive across the organization

ANALYTICAL

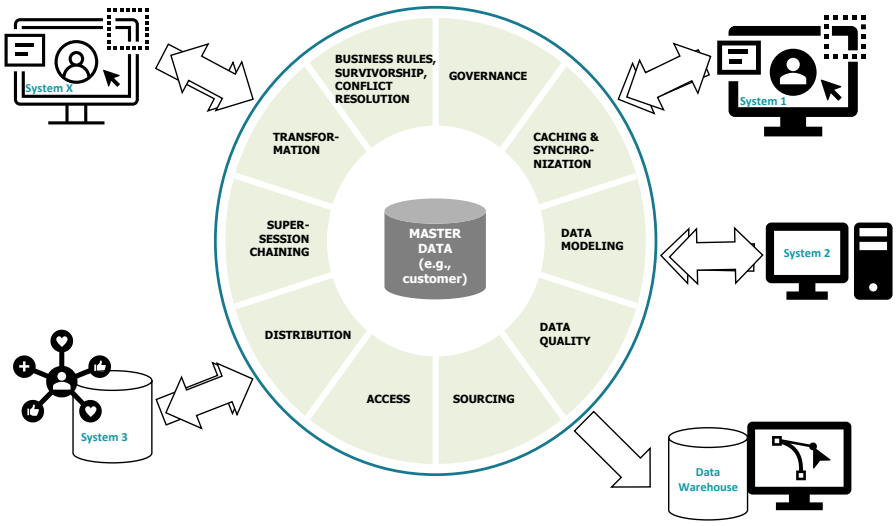
- MDM (usually a hub) created for use in Data Warehouse / Business Intelligence solutions
- Data integration (e.g., ETL) from source systems still required
- Data cleanse, consolidate, merge, de-duplicate, enrich etc. still required
- Master Data used for analytical purposes to ensure consistency of analyses, findings etc.
- Master Data NOT used by live operational line of business solutions, therefore operational systems do not see any benefits of MDM approach
- NOT essential to address business rules (such as survivorship, channels, governance) for operational MDM
- Easier (organizationally) to implement
- Impact of failure only on BI activities

DMA DATA MANAGEMENT ADVISORS

MDM Implementation

Analytical

Operational



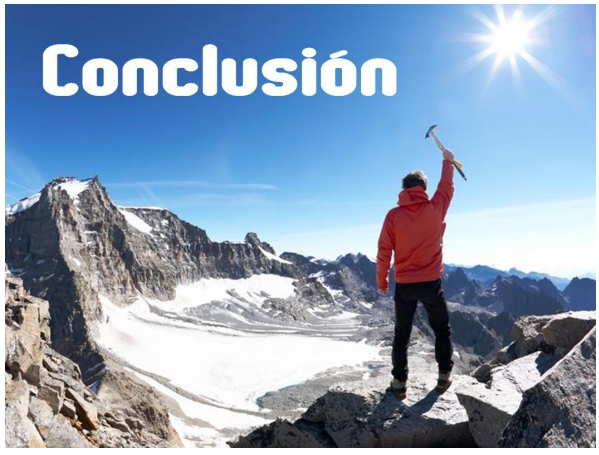
DMA DATA MANAGEMENT ADVISORS

It's NOT "The Field of Dreams"

- Projects cannot afford to wait for Master Data to be delivered
- Subsets of the Master Data must be delivered "just in time" as the Business projects need to use them
- This is NOT "The Field Of Dreams"
- This means **aligning** the **MDM initiatives** with the **Business Projects**

DMA
DATA MANAGEMENT ADVISORS

Conclusions



Don't believe you can *"build it and they will come"*

An MDM initiative needs to be just ahead of the business projects & deliver just in time aligned to the business initiatives.

Data models for MDM programs are essential

Determine data areas to master.

Configure MDM product.

Governance .. E.g. extended with rich metadata including Ownership, Source systems, Transformations etc.

Data Governance is vital

Operational MDM initiatives **cannot** be successful without addressing Data Governance for that data subject area.

Beware the hype on MDM technologies

A hub is NOT the only way.

Rarely will one MDM technology be appropriate / best for ALL Data Subject Areas.

DMA
DATA MANAGEMENT ADVISORS

Key Points (DMBoK 2 Rev)

- Shared Reference and Master data **belongs to the organization**, not to a particular application or department
- Reference and Master data management is an **on-going data quality improvement program**; its goals cannot be achieved by one project alone
- **Business data stewards** are the authorities **accountable** for controlling **reference data** values. Business data stewards work with data professionals to improve the quality of reference and master data
- Request, communicate, and, in some cases, approve of changes to reference data values before implementation
- Golden data values (**Golden records**) represent the organization's **best efforts** at determining the most accurate, current, and relevant data values for contextual use. New data may prove earlier assumptions to be false. Therefore, apply matching rules with caution, and ensure that any changes that are made are reversible
- Replicate master data values only from the **system of record**
- *A **System of Record** is an authoritative system where data is created/captured, and/or maintained through a defined set of rules and expectations (e.g., an ERP system may be the System of Record for sell-to customers).*
- ***A System of Reference** is an authoritative system where data consumers can obtain reliable data to support transactions and analysis, even if the information did not originate in the system of reference. MDM applications, Data Sharing Hubs, and Data Warehouses often serve as systems of reference.*

Links to Additional Content

- Berson, Alex and Larry Dubov. Master Data Management and Customer Data Integration for a Global Enterprise. McGraw-Hill, 2007. ISBN 0-072-26349-0. 400 pages.
- Brackett, Michael. Data Sharing Using A Common Data Architecture. New York: John Wiley & Sons, 1994. ISBN 0-471-30993-1. 478 pages.
- Chisholm, Malcolm. Managing Reference Data in Enterprise Databases: Binding Corporate Data to the Wider World. Morgan Kaufmann, 2000. ISBN 1-558-60697-1. 389 pages.
- Dreibelbis, Allen, Eberhard Hechler, Ivan Milman, Martin Oberhofer, Paul van Run, and Dan Wolfson. Enterprise Master Data Management: An SOA Approach to Managing Core Information. IBM Press, 2008. ISBN 978-0-13-236625-0. 617 pages.
- Dyche, Jill and Evan Levy. Customer Data Integration: Reaching a Single Version of the Truth. John Wiley & Sons, 2006. ISBN 0-471-91697-8. 320 pages.
- Finkelstein, Clive. Enterprise Architecture for Integration: Rapid Delivery Methods and Techniques. Artech House Mobile Communications Library, 2006. ISBN 1-580-53713-8. 546 pages.
- Loshin, David. Master Data Management. Morgan Kaufmann, 2008. ISBN 98-0-12374225-4. 274 pages.

Ref	Question	A	B	C	D	E
MDM1	What is a common motivation for Reference & Master Data Management?	The need to improve data quality and data integrity across multiple data sources	The need to build a Data Dictionary of all core data entities & attributes	Regulatory acts such as BCBS239, GDPR and SOX	The need to consolidate all data into one physical database	Business Intelligence & Data Warehousing
MDM2	Which of these is a valid definition of Master Data?	Data that if missing or incorrect will cause transactions and processes to fail	Data that is only held in one data source	Data that other data sits hierarchically beneath	Data that rarely, if ever, changes	Data about the business entities that provide context for business transactions
MDM3	Which of these is a valid definition of Reference Data?	Data that is fixed and never changes	Data used to classify or categorize other data	Data that provides metadata about other data entities	Data that is widely accessed and referenced across an organisation	Data that has a common and widely understood data definition
MDM4	Which of the following is NOT a primary Master Data Management area of focus?	Generating a golden record / best version of the truth	Identifying duplicate records	Producing read only versions of key data items	Providing access to golden data records	Producing clear data definitions for Master Data
MDM5	A strong argument for pursuing a Reference Data and/or Master Data management initiative is:	It will not require a lot of time or effort	They are essential functions in the data management framework	Job security for the data people	By centralizing the management of Reference and Master data, the organization can conform critical data needed for analysis	Application retirement
MDM6	A common driver for initiating a Reference Data Management program is:	It will improve data quality and facilitate analysis across the organization	It can be a one-time-only project	Managing codes and descriptions requires little effort and low cost	It will consolidate the process of securing third party code sets	Application simplification
MDM7	Reference Data Management includes defining relationships within and across domain value lists.	TRUE	FALSE			
MDM8	Which one of the following statements is true?	Business data stewards maintain lists of valid data values for master data instances.	Managing reference data requires the same activities and techniques as does managing master data.	Reference Data Management involves identifying the 'best' or 'golden' record for each domain.	Master Data Management requires techniques for splitting or merging an instance of a business entity.	Operational Master Data Management can be introduced before the Data is Governed
MDM9	An authoritative system where data is created/captured, and/or maintained through a defined set of rules and expectations is called	A System of Record.	A System of Origin.	A System of Referential Integrity.	A System of Retirement.	A System of Systems.

AFTER QUIZ 4

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)

Maximum possible score = 27

60% (CDMP Associate) = 16

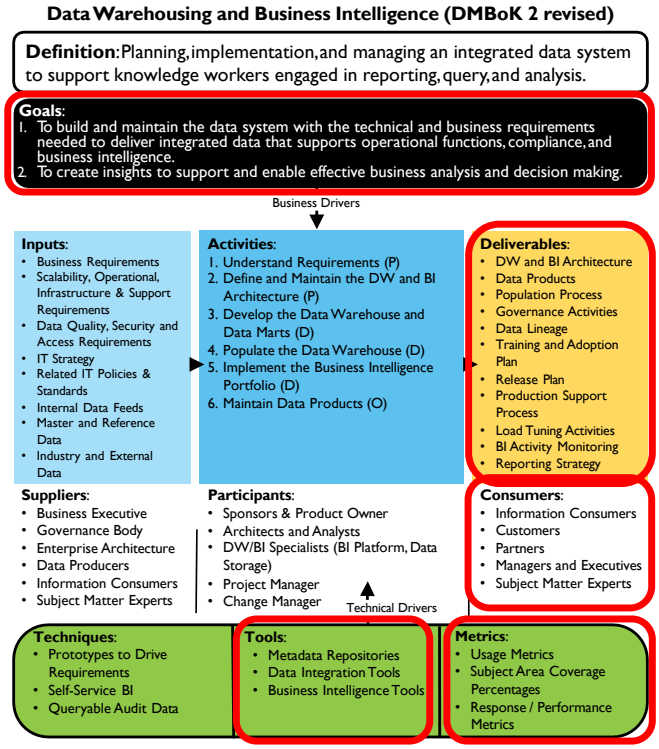
70% (CDMP Practitioner) = 19

80% (CDMP Master) = 22

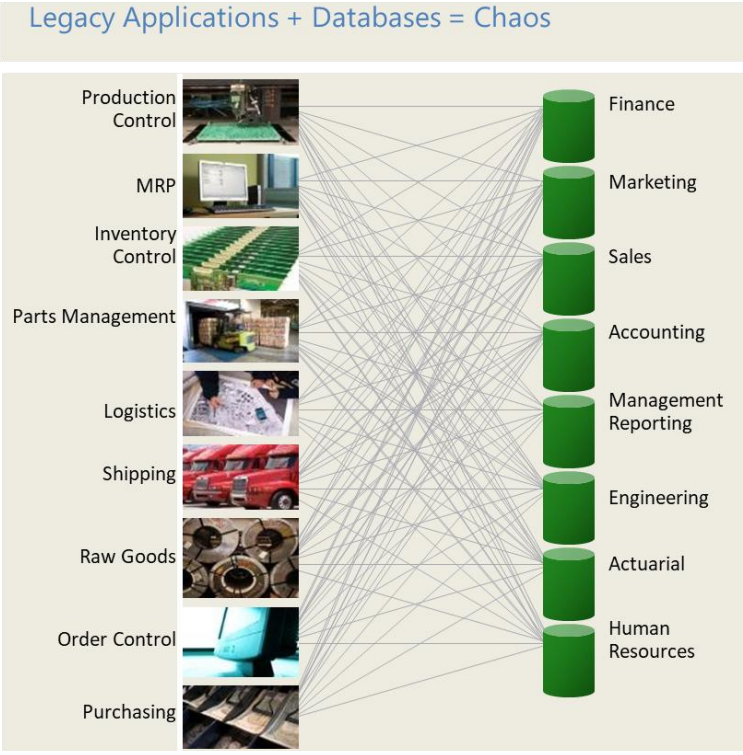
Data Warehousing & Business Intelligence

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA
DATA MANAGEMENT ADVISORS



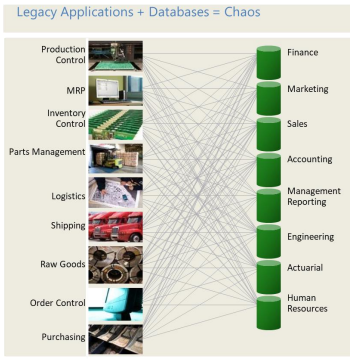
Why Use A Data Warehouse?

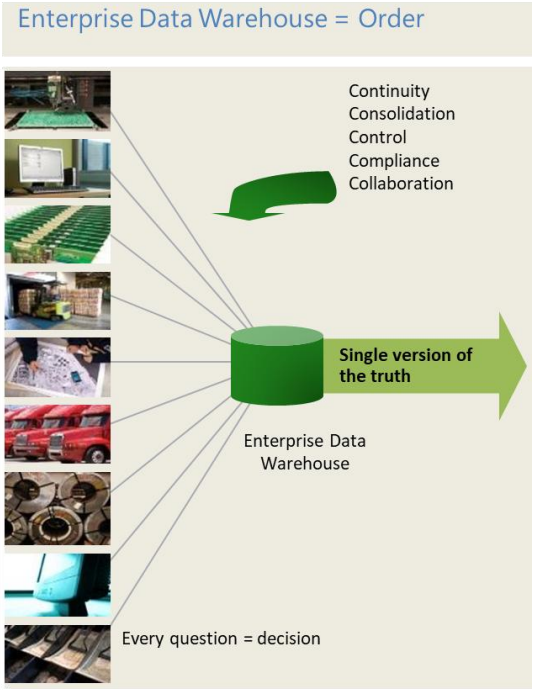


Why Use A Data Warehouse?

Purposes of a Data Warehouse:

1) Standardize data from multiple sources;
2) Save time building reports & analyses;
3) Report & analyze in ways you could not do before;
4) Prevent adverse impact on Operational systems;
5) Enable comparison without the underlying data changing;
6) ....





Classic Characteristics of a DW (Bill Inmon)

"a subject oriented, integrated, time variant and non-volatile collection of summary and detailed historical data used to support the strategic decision-making processes for the corporation"

Subject-Oriented:

A data warehouse can be used to analyze a particular subject area. For example, "sales" can be a particular subject.

Integrated:

A data warehouse integrates data from multiple data sources. For example, source A and source B may have different ways of identifying a product, but in a data warehouse, there will be only a single way of identifying a product

Time-Variant:

Historical data is kept in a data warehouse. For example, one can retrieve data from 3 months, 6 months, 12 months, or even older data from a data warehouse. This contrasts with a transactions system, where often only the most recent data is kept. For example, a transaction system may hold the most recent address of a customer, where a data warehouse can hold all addresses associated with a customer

Non-volatile:

Once data is in the data warehouse, it will not change. So, historical data in a data warehouse should never be altered.

Data Warehouse implementation guiding principles

Focus on business goals: Make sure DW serves organizational priorities and solves business problems.

Start with the end in mind: Let the business priority and scope of end-data-delivery in the BI space drive the creation of the DW content.

Think and design globally; act and build locally: Let end-vision guide the architecture, but build and deliver incrementally, through focused projects or sprints that enable more immediate return on investment.

Summarize and optimize last, not first: Build on the atomic data. Aggregate and summarize to meet requirements and ensure performance, not to replace the detail.

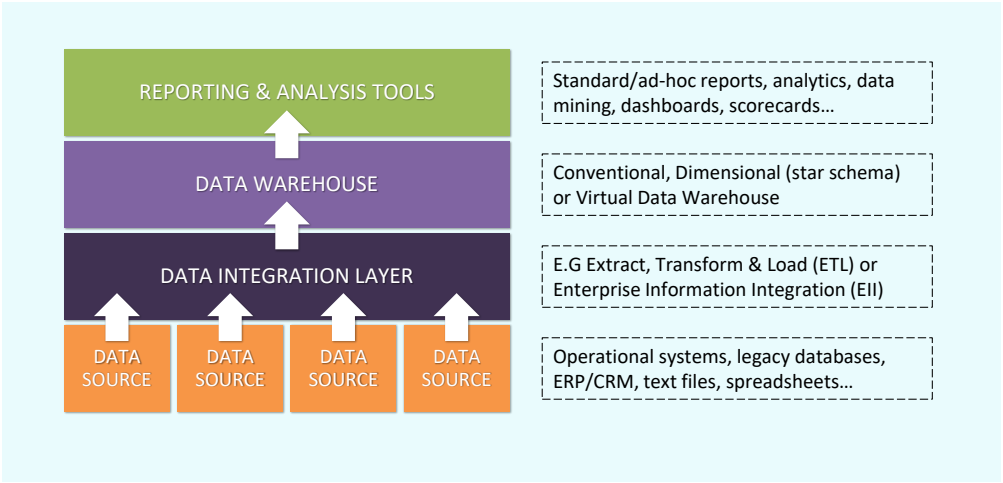
Promote transparency and self-service: The more context (Metadata of all kinds) provided; the better able data consumers will be to get value out of the data. Keep stakeholders informed about the data and the processes by which it is integrated.

Build Metadata with the warehouse: Critical to DW success is the ability to explain the data.
EG, "Why is this sum X?"
"How was that computed?",
"Where did the data come from?"
Metadata should be captured as part of the development cycle and managed as part of ongoing operations.

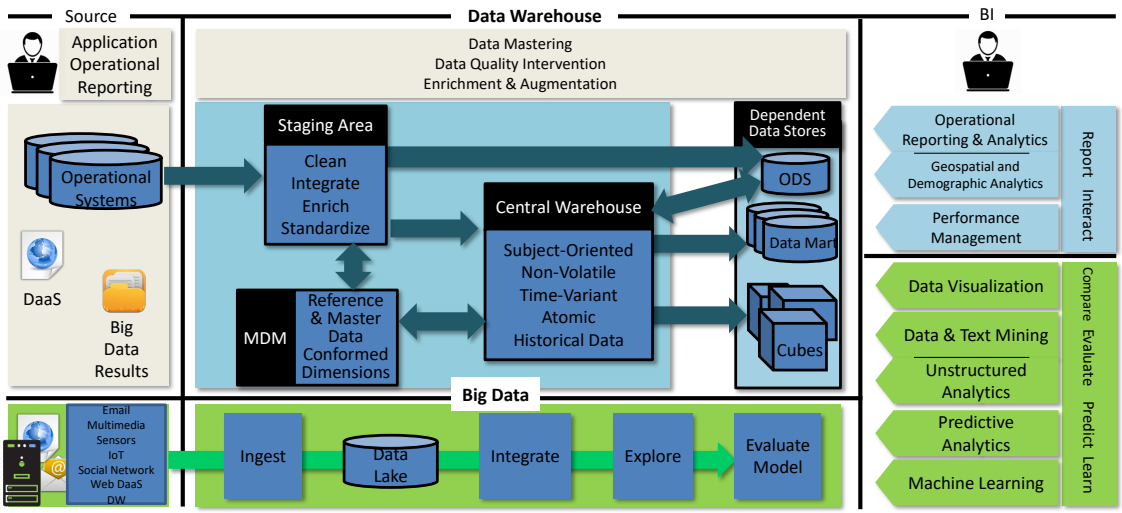
Collaborate: Collaborate with other data initiatives, especially those for Data Governance, Data Quality, and Metadata.

One size does not fit all: Use the right tools and products for each group of data consumers.

Simplified Business Intelligence Stack



Conceptual DW/BI and Big Data Architecture



	Data Lake	Data Warehouse	Data Mart
Data Scope	Broad, Raw Data	General, Cleaned	Focused
Prior Processing	None – Light	Moderate - High	Very High
Analysis Limitations	Only limited by input sources	Limited by Data cleaning / transformation choices	Limited to the Mart topic focus
Ease of Navigation	Difficult	Moderate	Easy

Types of Data Warehousing

CLASSIC

for Strategic and Tactical BI

- Data is loaded to the DW by a batch process; typically nightly
- All data in the DW is non-volatile
- Allows what-if analyses & re-testing of hypotheses against a stable data set

ACTIVE

for Strategic, Tactical and Operational BI

- Data, including volatile data, is loaded more frequently
- Drivers: lower latency and the need for more real-time or near real-time data
  - E.g. Real-time balances at a cash point (ATM)
- Isolation of change
  - Isolate volatile data from historical non-volatile data (e.g. put volatile data in an ODS)
  - Use union queries or Data Virtualisation to combine volatile and non-volatile data.
- Alternatives to batch ETL
  - Data Virtualisation
  - Message queues
  - Trickle-feeds / Pipelining
  - Event-based updates (Change Data Capture)

Big Data

Early definition: 3 V's

Volume:
The amount of data. Big data often has thousands of entities or elements in billions of records.

Velocity:
The speed at which data is captured, generated, or shared. Big data is often generated and can also be distributed and even analyzed in real time.

Variety:
The forms in which data is captured or delivered.

Expanded definition: 6/7 V's

Viscosity:
How difficult the data is to use or integrate.

Volatility:
How often data changes occur and therefore how long the data is useful.

Veracity:
How trustworthy the data is.

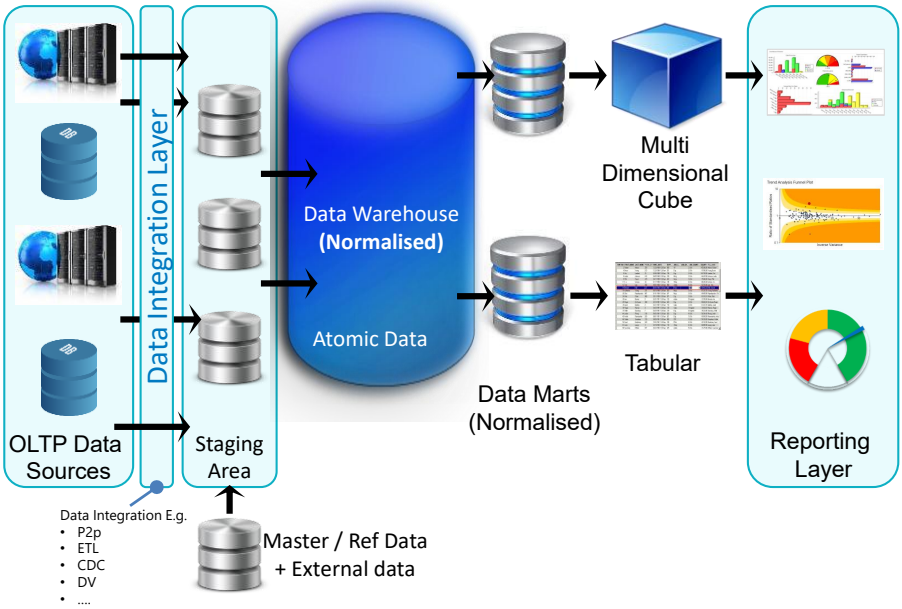
Variability:
Big Data requires storage of multiple formats; data structure is often inconsistent within or across data sets.

What is Data Warehousing? (DMBoK)

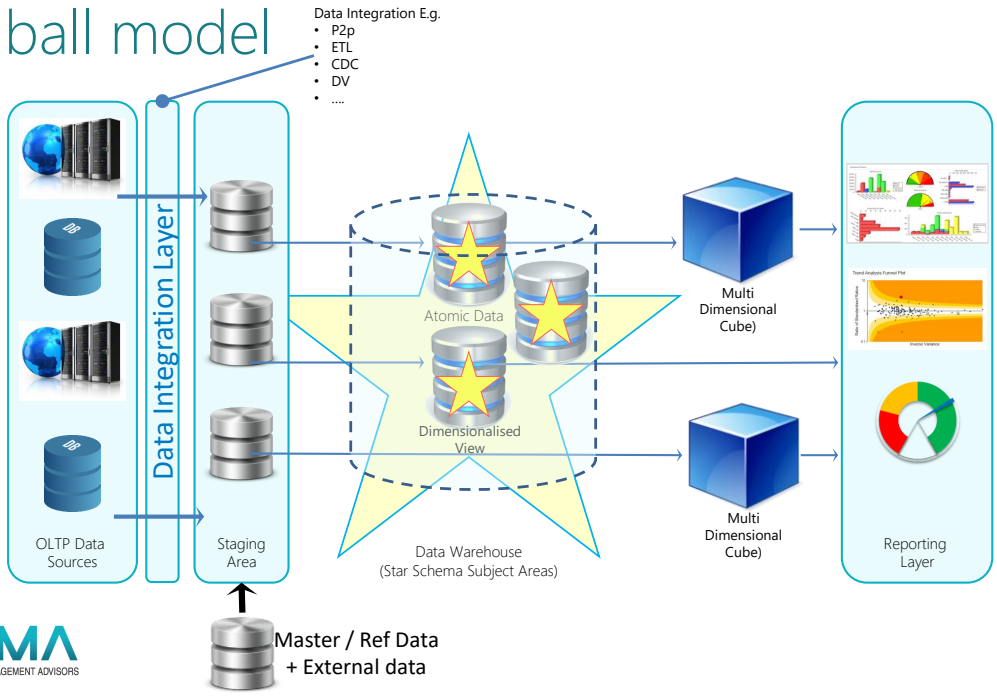
Data Warehousing is the term used to describe the processes that maintain the data contained within a data warehouse, namely:

- Extract processes
- Cleansing processes
- Transformation processes
- Load processes
- Associated Control processes
- The use of Meta-data

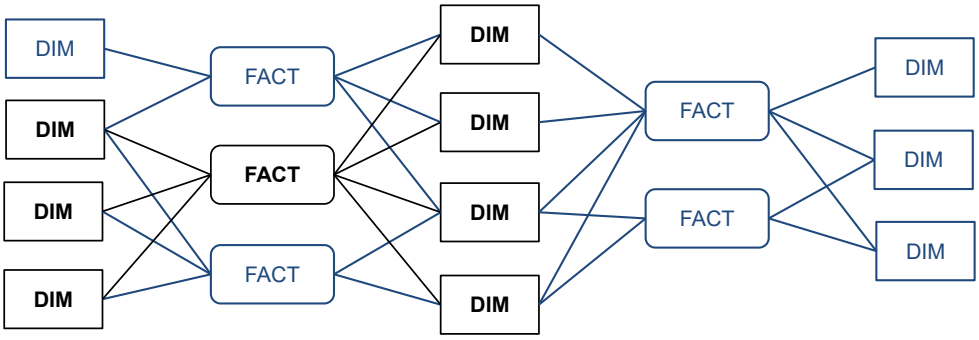
Inmon Data Warehouse Model



Kimball model



How do Dimensional Models fit into the Data Warehouse?



DMA
DATA MANAGEMENT ADVISORS

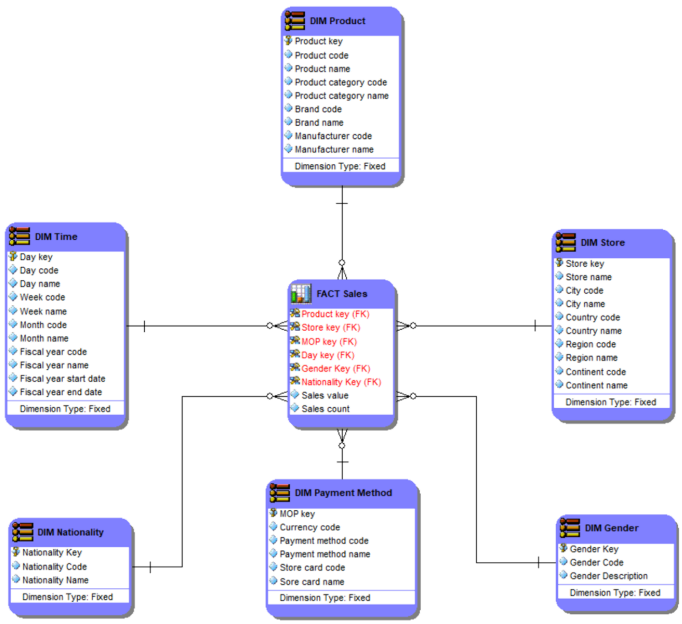
A Dimensional Model

Dimension tables

- Examples: Location, Product, Time, Promotion, Organisation etc.
- Records in the dimension tables correspond to nouns.
- The data in the dimension tables changes slowly – the number of new records created each day is typically low.

Fact tables

- Contains measures (e.g. Sales Value GBP) and dimension columns
- Records in the fact tables correspond to events, transactions, or measurements.
- The number of new records created each day is typically high.



DMA
DATA MANAGEMENT ADVISORS

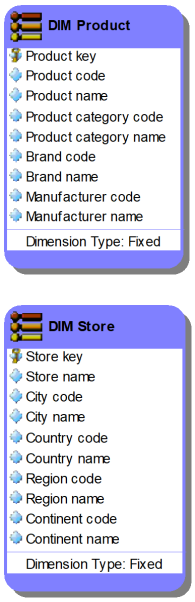
Dimensions & Hierarchies

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 redundancy



DMA
DATA MANAGEMENT ADVISORS

Fact Tables

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

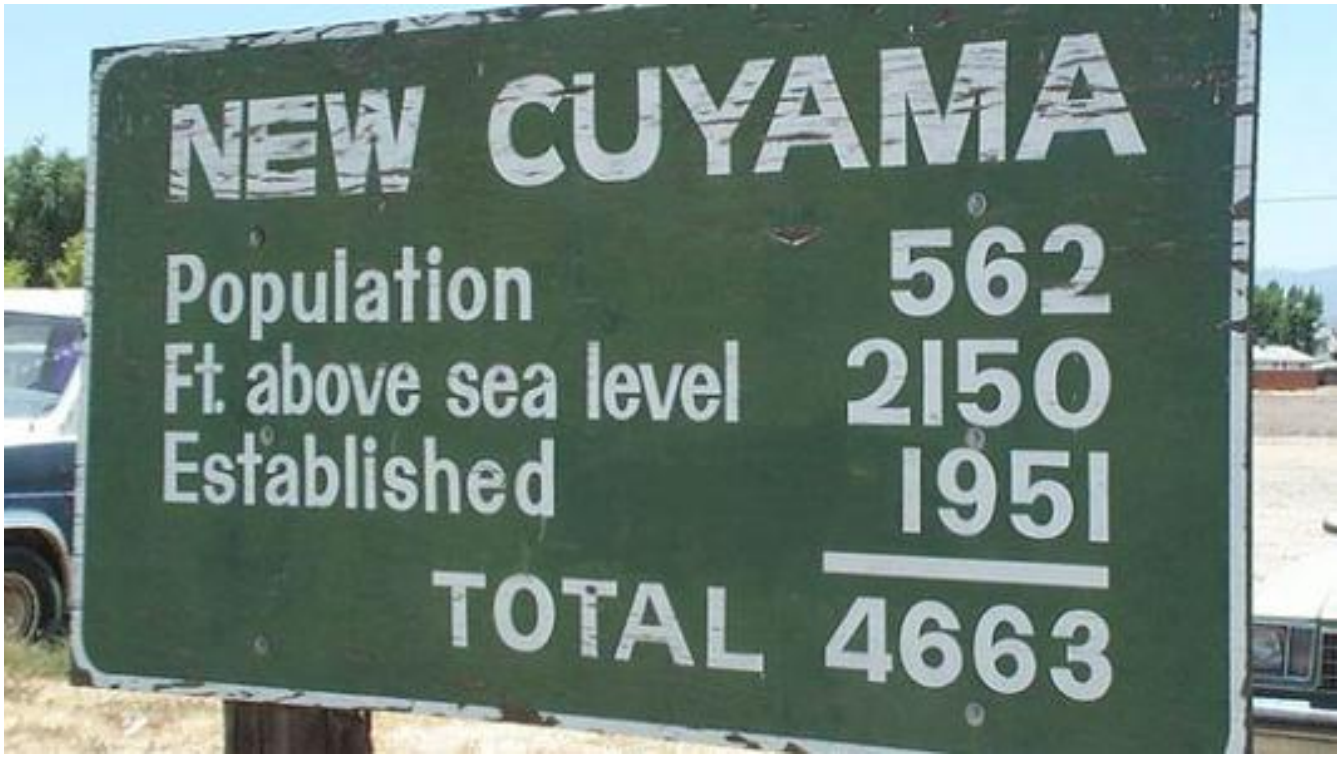
Data is added regularly; the tables are "long" – often millions of records

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

Low redundancy

The most useful measures are "additive"

DMA
DATA MANAGEMENT ADVISORS



How can we handle slowly changing dimensions?

There are standard techniques for handling slowly changing dimensions.

- Type 1 (overwrite)
- Type 2 (add new row)
- Type 3 (add new attribute)
- Type 4 (add history table)
- Type 6 (hybrid)
- Others – see the internet!

We may need to employ different techniques for different fields.

DMA
DATA MANAGEMENT ADVISORS

Types of Business Intelligence Tool

Query and Reporting Tools	• Enable business users to create ad hoc queries and reports. Also used to create standard reports. • Examples: Business Objects, PowerBI, Cognos, Crystal Reports
OLAP Tools	• Support interactive multi-dimensional analysis • Example: Hyperion, MS Analysis Services with Excel
Analytic Applications	• "Industry Standard Data Marts in a Box" • Include logic and processes to extract data from well-known source systems (e.g. vendor ERP systems) • Also include a data model for the data mart and pre-built reports and dashboards.
Management Dashboards and Scorecards	• Dashboard = dynamic presentation of operational information • Scorecard = static representation of progress towards longer term goals
Performance Management Tools	• Tools to monitor performance (typically of the organisation) against goals. Such tools use data from source systems but also capture data directly, e.g., about goals and activities. • Applications include budgeting, planning and financial consolidation
Predictive Analytics and Data Mining	• Data Mining - tools that allow users to discover patterns, often using various algorithms, by exploring data interactively, without having a specific question to start with. • Predictive analytics – tools that predict the future and support "what if" analysis.
Advanced Visualisation and Discovery Tools	• Use an in-memory architecture to allow users to interact with the data in a highly visual, interactive way. • Example: PowerBI, Tableaux, QlikView

Summary

- A Data Warehouse is a technology solution that addresses the problem of how to support **business intelligence** needs.
- Data Warehouses contain data from source systems together with externally provided data that is combined & stored in a manner optimised for **query intensive use**
- Data Warehouses must contain **integrated** data in order to be more than just a warehouse of data.
- Dimensional Data Modelling is used to make data **easy to use** and to support **fast queries**.
- Data Warehousing processes must be transparent (open) and visible in order for business users to **understand** and have **confidence** in data.
- **Many** types of **Business Intelligence tools** exist

Ref	Question	A	B	C	D	E
DW1	Which of the following uses for a Data Mining tool is not optimal?	Identification of data quality issues with your SAP Financial system	Fraud Detection	Customer Segmentation and Scoring	Predictive Analysis	Identifying potential loan defaulter's
DW2	Which of the following is not a good example of BI?	Strategic Analytics for Business Decisions	Decision Support Systems	Supporting Risk Management Decision Reporting	Statutory reporting to a Regulatory Body	Identifying top quartile customers
DW3	According to Henry Morris of IDC, Analytic Applications provide business with a pre-built solution to optimize a functional area or industry segment	TRUE	FALSE			
DW4	When performing an evaluation of analytic applications, which of the following questions is least relevant to identify the level of effort needed?	How much of the tool infrastructure meets our organisational infrastructure	The Standard source systems for which ETL is supplied	No. of source systems we need to integrate into the tool	How much do the canned processes in the tool match our business	Annual costs such as license, maintenance, etc
DW5	You need to discover possible relationships or to show data patterns in an exploratory fashion when you do not necessarily have a specific question to ask. What kind of data tool would you use to identify patterns of data using various algorithms?	ETL Jobs	Data Quality Profiler	Meta-Data Data Lineage View	Data Mining	Data Visualisation Application
DW6	"Slice", "Dice", "Roll-up" and "Pivot" are terms used in what kind of data processing?	OLAP	OLTP	ODS	EDI	EIEIO
DW7	A comparatively new architectural approach is where volatile data is provisioned in a data warehouse structure to provide transactional systems with a combination of historical and near real time data to meet customer needs. This is a definition of:	Operational Data Store	Behavioural Decision Support Systems	Active Data Warehousing	On Line Transactional Processing System	On Line Analytical Processing Cube

AFTER QUIZ 5

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)

Maximum possible score = 34

60% (CDMP Associate) = 21

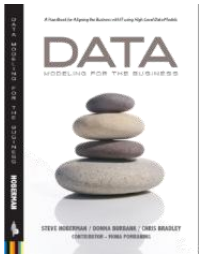
70% (CDMP Practitioner) = 24

80% (CDMP Master) = 28

Data Modelling & Design

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA DATA MANAGEMENT ADVISORS



"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

Data Modeling and Design **DMBoK2 Revised**

Definition: Data modeling is the process of discovering, analyzing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative and involves conceptual, logical, and physical models

Goals:
1. To confirm and document an understanding of different perspectives, which leads to applications that more closely align with current and future business requirements.
2. To understand how data fits together and creates a foundation to successfully complete broad-scoped initiatives such as master data management and data governance.
3. To lower support costs and reduce the costs of building new applications thanks to increased reusability opportunities.

Business Drivers

Inputs:
- Enterprise Data Model
- Existing Data Models and Databases
- Data Standards
- Data Sets
- Data Requirements
- Data Architecture
- Enterprise Taxonomy

Activities:
1. **Plan for Data Modeling (P)**
2. **Build the Data Model (D)**
   1. Conceptual Data Modeling
   2. Logical Data Modeling
   3. Physical Data Modeling
3. **Review the Data Models (C)**
4. **Maintain the Data Models (O)**

Deliverables:
- Diagrams (different levels, schemes, and notations)
- Definitions
- Issues and outstanding questions
- Lineage

(P) Planning,
(C) Control,
(D) Development,
(O) Operations

Suppliers:
- Business Professionals
- Business Analysts
- Data Architects
- Database Administrators and Developers
- Subject Matter Experts
- Data Stewards
- Metadata Administrators

Participants:
- Business Analysts
- Data Analysts
- Data Modelers
- Data Consumers

Consumers:
- Business Analysts
- Data Modelers
- Database Administrators and Developers
- Software Developers
- Data Stewards
- Data Quality Analysts
- Data Consumers

Technical Drivers

Techniques:
- Naming conventions
- Forward / reverse engineering
- Approach determination

Tools:
- Data modeling tools
- Lineage tools
- Data profiling tools
- Metadata repositories
- Data model patterns
- Industry data models

Metrics:
- Data model scorecard

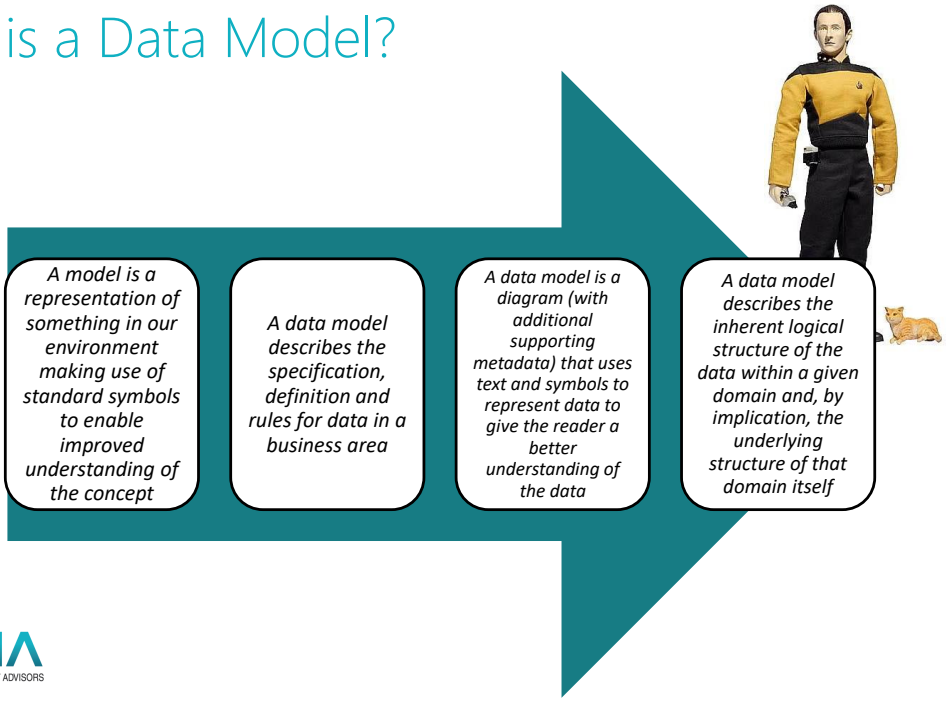
DMA DATA MANAGEMENT ADVISORS

The deliverables of the data modeling process include:

- **Diagram**: A data model contains one or more diagrams. The diagram is the visual that captures the requirements in a precise form. It depicts a level of detail (e.g., conceptual, logical, or physical), a scheme (relational, dimensional, object-oriented, fact-based, time-based, or NoSQL), and a notation within that scheme (e.g., information engineering, unified modeling language, object-role modeling).
- **Definitions**: Definitions for entities, attributes, and relationships are essential to maintaining the precision on a data model.
- **Issues and outstanding questions**: Frequently the data modeling process raises issues and questions that may not be addressed during the data modeling phase. In addition, often the people or groups responsible for resolving these issues or answering these questions reside outside of the group building the data model. Therefore, often a document is delivered that contains the current set of issues and outstanding questions. An example of an outstanding issue for the student model might be, "If a **Student** leaves and then returns, are they assigned a different **Student Number** or do they keep their original **Student Number**?"
- **Lineage**: For physical and sometimes logical data models, it is important to know the data lineage, that is, where the data comes from. Often lineage takes the form of a source/target mapping, where one can capture the source system attributes and how they populate the target system attributes. Lineage can also trace the data modeling components from conceptual to logical to physical within the same modeling effort. There are two reasons why lineage is important to capture during the data modeling. First, the data modeler will obtain a very strong understanding of the data requirements and therefore is in the best position to determine the source attributes. Second, determining the source attributes can be an effective tool to validate the accuracy of the model and the mapping (i.e., a reality check).

Page 149 / 152 in printed version
Page 156-157 of DM-BOK pdf
Section 2.1 (Activities / Plan for Data Modelling)

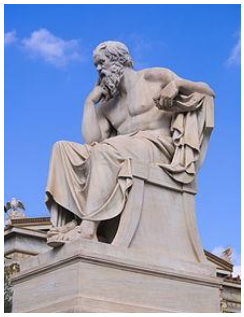
What is a Data Model?



DMA
DATA MANAGEMENT ADVISORS

Clarity & Definition is vital

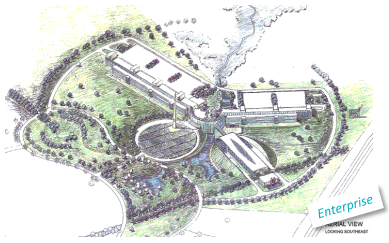
"The beginning of wisdom is the definition of terms."

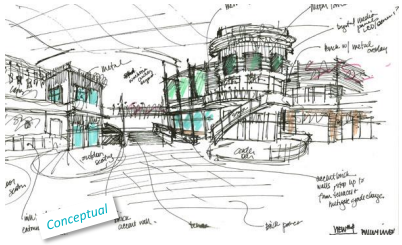


Socrates: 470 – 399 BC

DMA DATA MANAGEMENT ADVISORS

We All Use Models



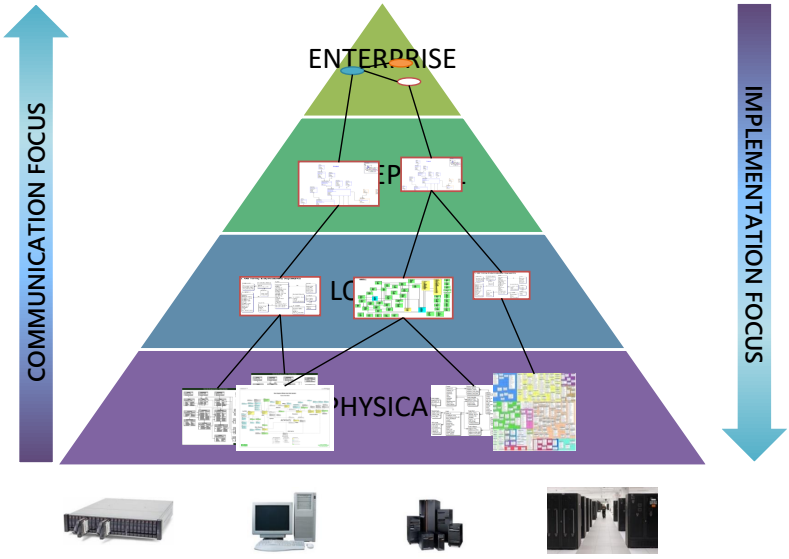


DM DATA MANAGEMENT ADVISORS

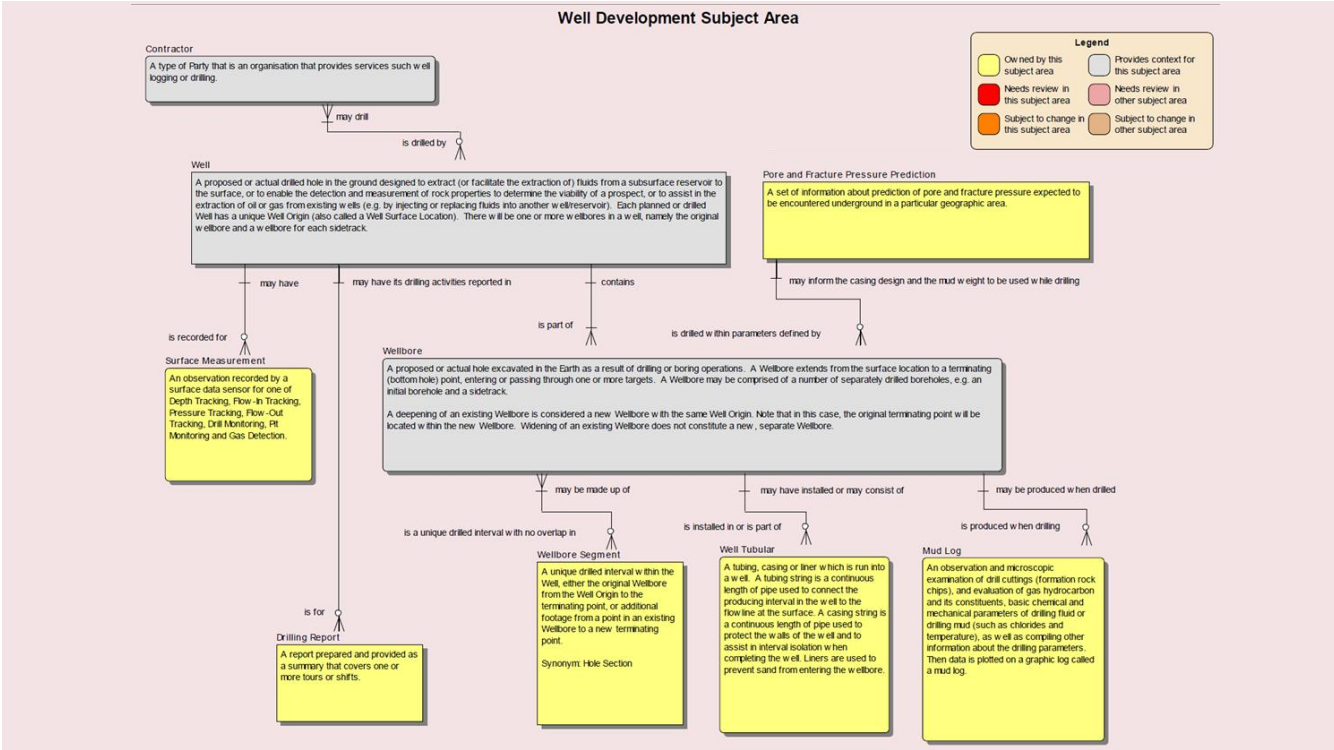
Data Model Levels

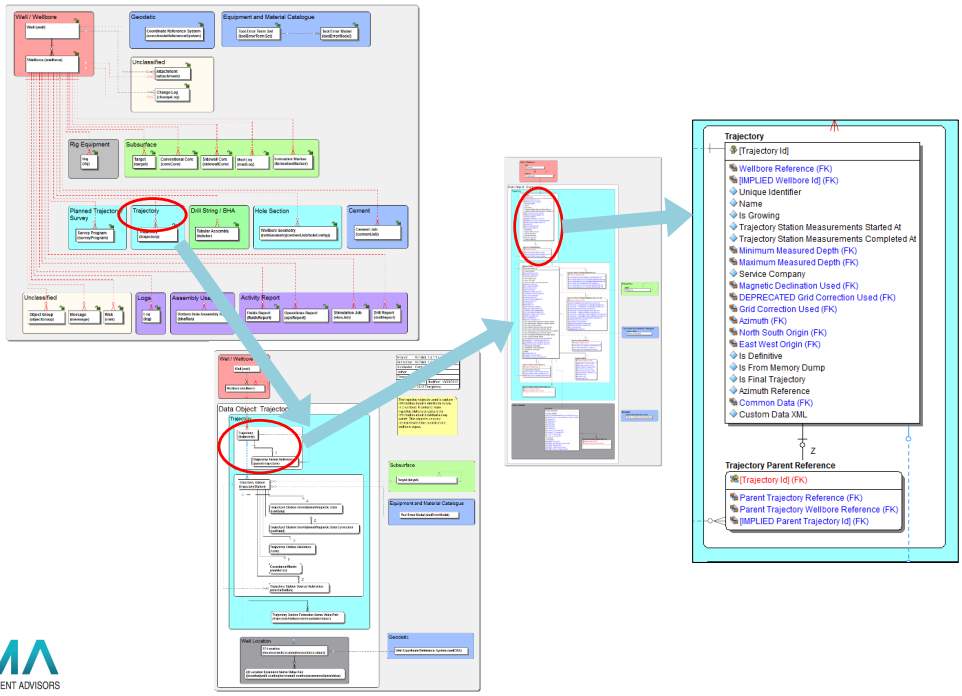
ANSI Standards Planning and Requirements Committee (SPARC) call these levels:

- *Conceptual,*
- *External, and*
- *Internal*



DMA
DATA MANAGEMENT ADVISORS





DMA
DATA MANAGEMENT ADVISORS

DMA
DATA MANAGEMENT ADVISORS

The main purpose of a data model is not to design a database, but to describe a business (C. Bradley CDMP Fellow 2010)

DMA DATA MANAGEMENT ADVISORS

A Data Model represents

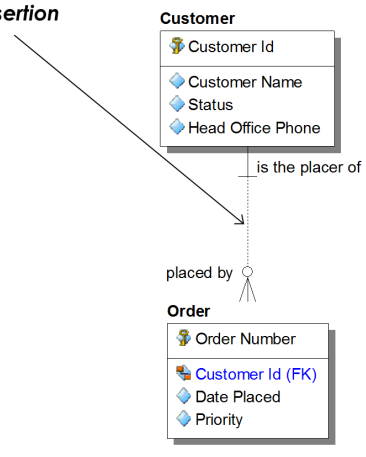
Classes of ***ENTITIES*** ***(kinds of things)*** *about which a company wishes to know or hold information*

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

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

The model describes the organization of the data ***irrespective*** *of how data might be represented in a computer system*

Relationships form a concrete Business Assertion



DMA DATA MANAGEMENT ADVISORS

A Data Model Represents (Entity Categories)

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

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, …*

DMA
DATA MANAGEMENT ADVISORS

Types of data modelled (DMBoK)

Category information:

Data used to classify and assign types to things. For example, customers classified by market categories or business sectors; products classified by color, model, size, etc. AKA Reference Data.

Resource information:

Basic profiles of key resources required to conduct operational processes such as Product, Customer, Supplier, Facility, Organization, and Account. AKA Master & Reference Data.

Business event information:

Data created while operational processes are in progress. Examples include Customer Orders, Supplier Invoices, Cash Withdrawal, and Business Meetings. AKA transactional business data.

Detail transaction information:

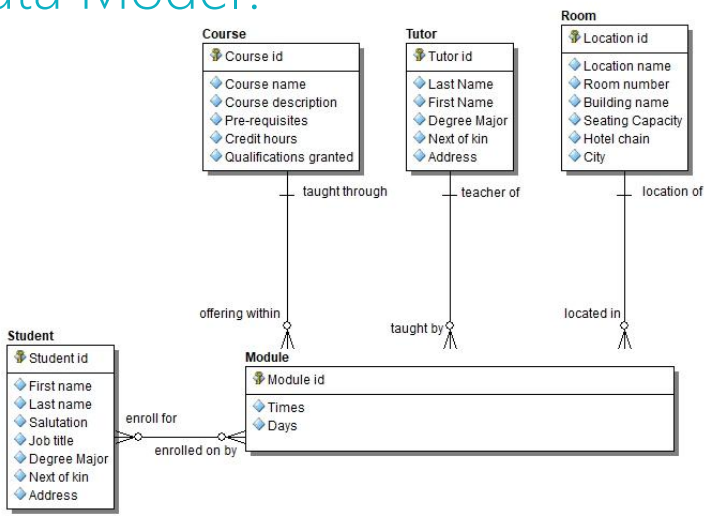
Often produced through point-of sale systems (either in stores or online). Also produced through social media systems, other Internet interactions (clickstream, etc.), and by sensors in IOT / machines (GPS, RFID, Wi-Fi, etc.). Often it is aggregated, used to derive other data, and analyzed for trends. This type of data (large volume and/or rapidly changing) is often referred to as Big Data.

DMA
DATA MANAGEMENT ADVISORS

What Is A Business 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

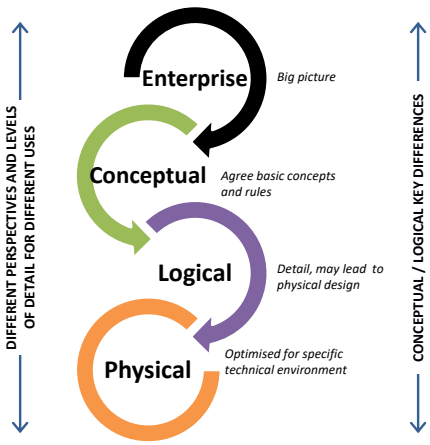


Business Assertions
- A Student enrolls for zero one or more modules
- A Course can be taught through zero one or more Modules
- A Room can be the location of zero one or more modules
- A Tutor can be the teacher of zero one or more modules

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?

Enterprise vs. Conceptual vs. Logical

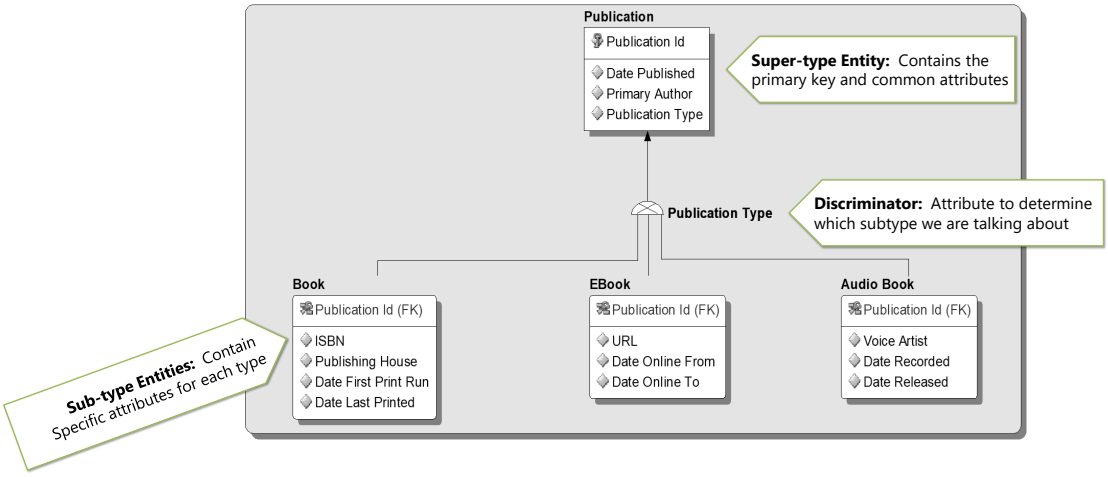


- Common understanding before progressing too far into detail
- Used to communicate with the Business
- Overview: main entities, super types, attributes, and relationships
- Lots of Many to Many & multi meaning relationships
- Relationships frequently show multiplicity of meaning
- May be denormalised
- Non-atomic & multi-valued attributes allowed; no keys
- Should fit on one page
- 20% of the modelling effort

CONCEPTUAL / LOGICAL KEY DIFFERENCES

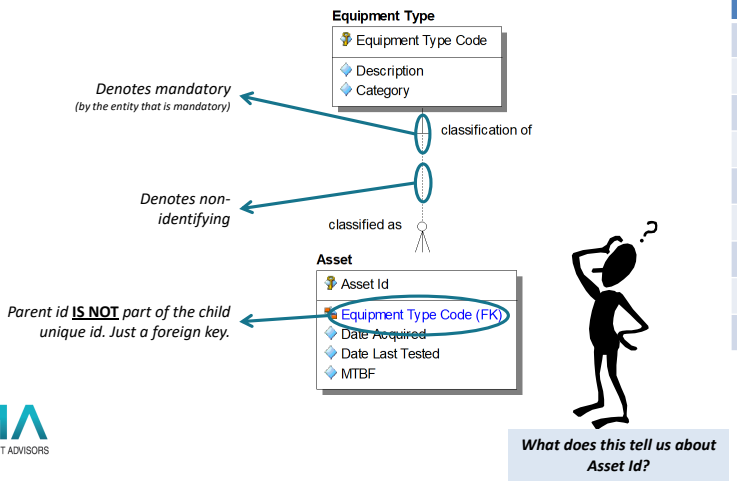
- Detailed: ~ 5x Entities vs Conceptual model
- Detailed: Frequently pre-cursor to 1st cut physical (database) design
- Detailed: Key input to requirements specification
- M:M relationships resolved: Intersection entities mostly have meaning
- Supertypes AND Subtypes included
- Relationship optionality added
- Primary, foreign, alternate keys included
- Reference entities included
- Fully normalized – no multi-valued, redundant, non-atomic attributes
- May be partitioned (sub-models)
- 80% of the modelling effort

Entity Subtypes



Relationship Types

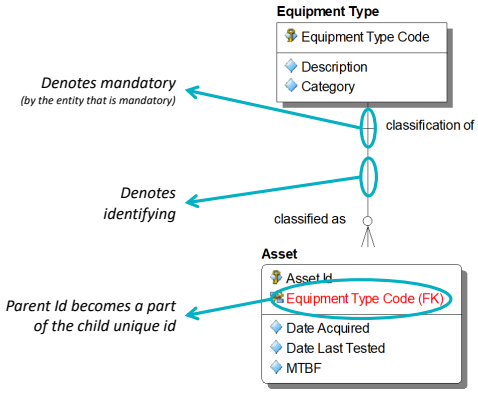
Non-Identifying



Eq Type	Eq Type Code	Asset Id
Laptop	L	1
Laptop	L	2
Laptop	L	3
Projector	P	4
Projector	P	5
Desktop	D	6
Desktop	D	7
Desktop	D	8
Desktop	D	9

Relationship Types

Identifying

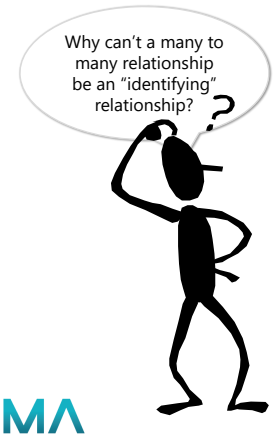


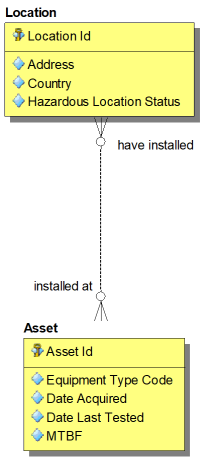
Eq Type	Eq Type Code	Asset Id
Laptop	L	1
Laptop	L	2
Laptop	L	3
Projector	P	1
Projector	P	2
Desktop	D	1
Desktop	D	2
Desktop	D	3
Desktop	D	4

What does this tell us about Asset Id?

Relationship Types

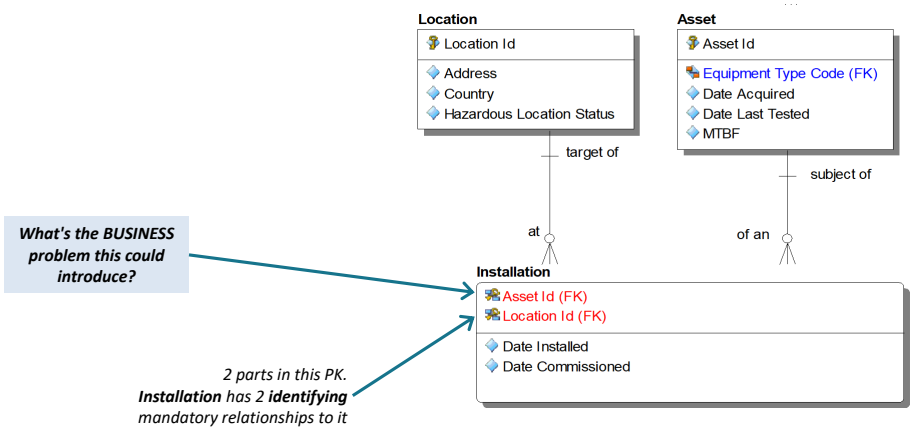
Many to Many *(AKA nonspecific)*





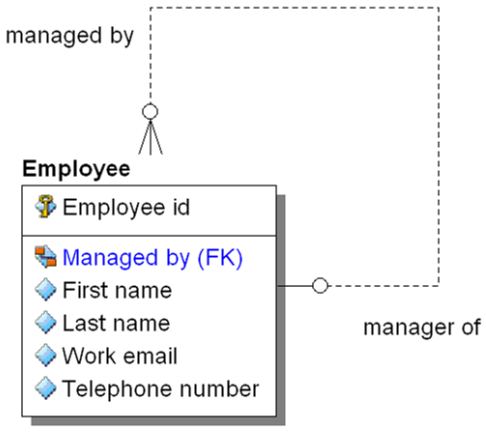
Chris's law
99% of M:M relationships represent a real business concept that is the intersection entity

Relationship Types

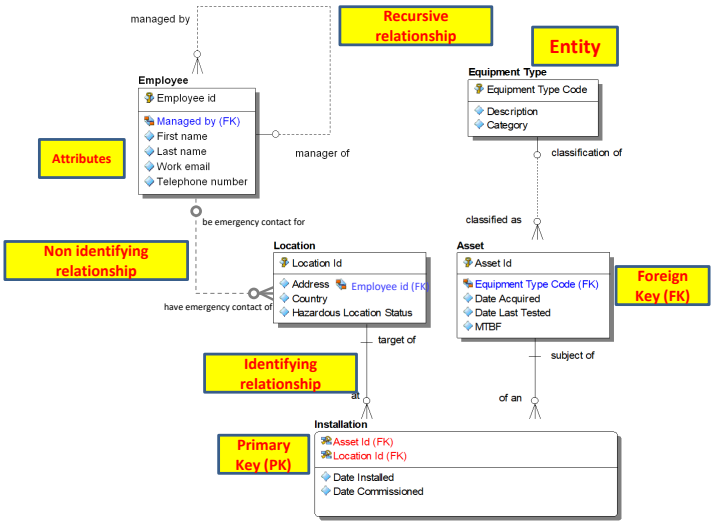


Recursive Relationships

- A recursive relationship occurs when there is a relationship **between an entity and itself**.
- E.g. A one-to-many recursive relationship occurs when an **employee** is the **manager** of other **employees**.
- The **employee** entity is related to itself, and there is a one-to-many relationship between one **employee** (the manager) and many other **employees** (the people who report to the manager).
- **But we cannot have duplicated attribute names in an entity; hence the FK is given a role name (e.g. "managed by").**



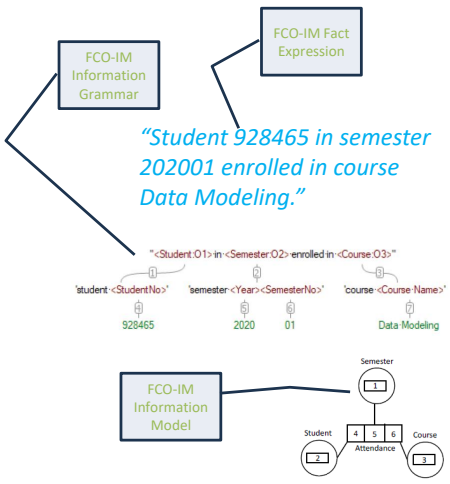
Logical Data Model Components



DMA
DATA MANAGEMENT ADVISORS

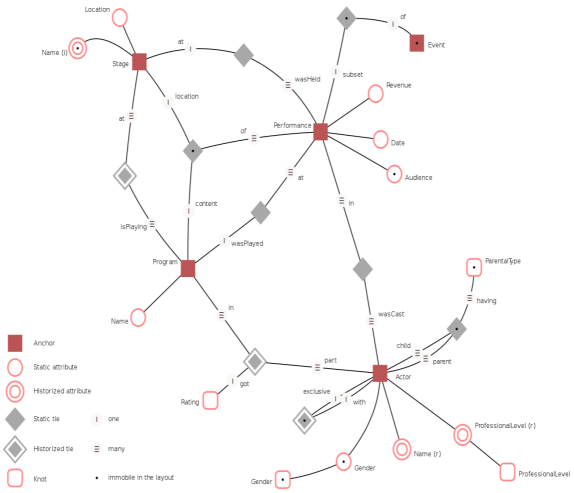
Overview of Data Modelling Notations

- **Entity-Relationship (ER) Notation:** For example, Information Engineering (IE), uses a set of symbols for representing entities, their attributes, and relationships
- **Barker Notation:** Developed by Richard Barker, used in the Oracle Designer CASE tool and is popular in Europe
- **IDEF1X Notation:** Used in the integration definition for information modelling, a popular Federal standard for data modelling
- **Fully Communication Oriented Modeling (FCO-IM):** The conceptual information models are made using natural language with concrete examples. Using so called elementary fact expressions, Information grammar & Information model; FCO-IM provides extensive support for different ways of communicating similar facts, synonyms, homonyms, generalized object types, and other conceptual modeling aspects.
- **UML Class Diagram Notation:** Used in the Unified Modelling Language for representing object-oriented software systems
- **Data vault modelling:** A database modelling technique that provides historical storage of data from multiple operational systems.



DMA
DATA MANAGEMENT ADVISORS

Data Vault & Anchor Models



Well suited to handle information that changes in structure and context over time.

Hubs, Satellites, and Links are the 3 entity types to focus its DV design around functional areas of business.

Data Vault Modeling is a hybrid approach, encompassing the best of breed between 3NF and Dimensional (star schema)

Anchors are shown as squares

DMA
DATA MANAGEMENT ADVISORS

Keys

DMA
DATA MANAGEMENT ADVISORS

Key Vocabularies

Terms used conceptually

- Primary key
- Alternate key
- Composite key
- Super key
- Candidate key
- Surrogate key
- …

Terms used physically

- Primary key
- Clustered key
- Encryption key
- Partitioning key
- Index
- Identity
- Sequence
- …

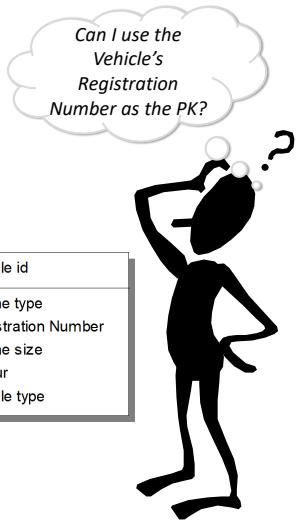
Ref. Karen Lopez, Information Advisors

Primary Key

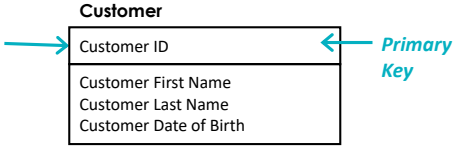
Mandatory

Must Uniquely Identify Each Instance Of The Entity

Unchanging



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



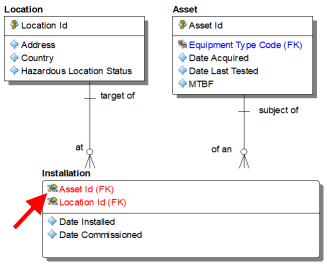
Composite vs Compound (DMBoK2 R)

A **composite key** is a candidate key that consists of two or more attributes that together uniquely identify an entity.

Student	Date	Period	Present
J. Smith	01/06/09	1	Y
B. Rogers	01/06/09	1	Y
A. Black	01/06/09	2	N
H. Rose	02/06/09	1	Y
M. Wright	02/06/09	2	N

These three fields are combined to make a composite primary key

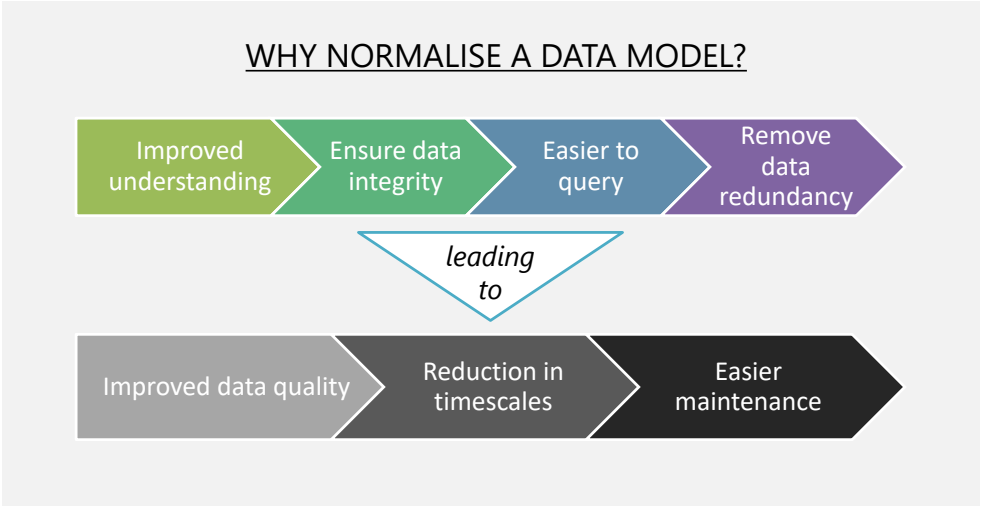
A **compound key** is a composite key for which each attribute that makes up the key is a foreign key in its own right



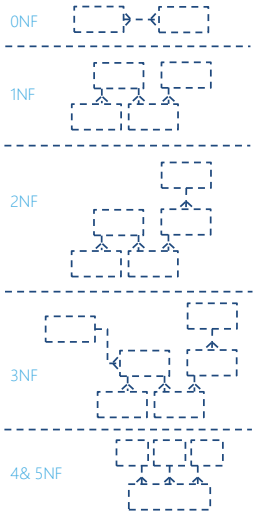
These 2 foreign keys make up the primary key of the entity and are a Compound primary key

A compound key is similar to a composite key in that two or more attributes are needed to create a unique value. However, a compound key is created when two or more primary keys from different entities are present as foreign keys within an entity. The foreign keys are used together to uniquely identify the entity.

Normalisation



Summary



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

Definition

1NF Every non-key attribute in an entity must depend on it's primary key

2NF Each entity must have the fewest possible correct primary key attributes

3NF Each non key element must be directly dependent upon the primary key and not upon any other non-key attributes

DMA
DATA MANAGEMENT ADVISORS

Physical database design best-practice

- Use normalised design for relational databases supporting OLTP apps.
- Use views, functions and stored procedures to create non-normalised, application-specific, object-friendly, conceptual (virtual) views of data.
- Use standard naming conventions.
- Enforce data security and integrity at the database level, not in the application.
- Keep database processing on the database server as much as possible.
- Grant permissions on database objects only to application groups or roles, not to individuals.
- Do not permit any direct, ad-hoc updating of the database.

DMA
DATA MANAGEMENT ADVISORS

Class, Prime, Modifier, Qualifier Words

The following word classification types are used by various data modelling tools and are defined below with examples.

Defining Word Classification Types

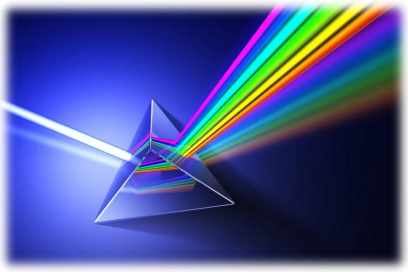
Prime Word:
The prime word identifies the object or element being defined. Typically, these objects represent a person, place, thing, or event about which an organization wishes to maintain information. Prime words may act as primary search identifiers when querying a database system and provide a basic list of keywords for developing a general-to-specific classification scheme based on business usages. ***CUSTOMER*** in ***Customer Address*** is an example of a ***prime word.***

Modifier:
A modifier gives additional information about the class word or prime word. Modifiers may be adjectives or nouns. ***DELIVERY*** in ***Customer Delivery Address*** is an example of a ***modifier***. *Other modifier examples: ANNUAL, QUARTERLY, MOST, and LEAST.*

Class Word:
A class word is the most important noun in a data element name. Class words identify the use or purpose of a data element. Class words designate the type of information maintained about the object (prime word) of the data element name. ***ADDRESS*** in ***Customer Address*** is an example of a ***class word.***

Qualifier:
A qualifier is a special kind of modifier that is used with a class word to further describes a characteristic of the class word within a domain of values, or to specify a type of information that can be attached to an object.
Examples: FEET, METERS, SECONDS, and WEEKS.

Physical database design best practices



Performance and Ease of Use

Ensure quick and easy access to data

Reusability

Multiple applications can use the data

Integrity

The data should have valid business meaning and value

Security

Data should only be available to authorised users

Maintainability

Ensure cost of maintenance does not exceed its value to the organisation

Transforming from a logical to physical data model

- **Denormalisation**
  Selectively and justifiably violating normalisation rules to reduce retrieval time, potentially at the expense of additional space, insert/update time and reduced data quality.
- **Surrogate keys**
  Substitute keys not visible to the business.
- **Indexing**
  Create additional index files to optimise specific types of queries.
- **Partitioning**
  Break a table or file horizontally or vertically.
- **Views**
  Virtual tables used to simplify queries, control data access and rename columns.
- **Dimensionality**
  Creation of fact tables with associated dimension tables. Structured as star schemas and snowflake schemas for BI.

DMA
DATA MANAGEMENT ADVISORS

Partitioning

Horizontal partitioning

Horizontal partitioning is the partitioning of a table into a number of smaller tables on the basis of rows. For example, in an employee table, employees with a salary of less than £25, 000 will be partitioned into a different table.

Vertical partitioning

Vertical partitioning is dividing the table based on the different columns. For example, in a customer table, retrieving only the name and contact number of customers into a different table.

DMA
DATA MANAGEMENT ADVISORS

ACID Test For Transaction Processing

Atomicity

Atomicity requires that database modifications must follow an "all or nothing" rule. Each transaction is said to be atomic. If one part of the transaction fails, the entire transaction fails and the database state is left unchanged.

To be compliant with the 'A', a system must guarantee the atomicity in each and every situation, including power failures / errors / crashes.

This guarantees that 'an incomplete transaction' cannot exist.

Consistency

The consistency property ensures that any transaction the database performs will take it from one consistent state to another.

Consistency states that only consistent (valid according to all the rules defined) data will be written to the database.

Quite simply, whatever rows will be affected by the transaction will remain consistent with each and every rule that is applied to them (including but not only: constraints, cascades, triggers).

While this is extremely simple and clear, it's worth noting that this consistency requirement applies to everything changed by the transaction, without any limit (including triggers firing other triggers launching cascades that eventually fire other triggers etc.) at all.

Isolation

The requirement that no transaction should be able to interfere with another transaction at all. In other words, it should not be possible that two transactions affect the same rows run concurrently, as the outcome would be unpredicted, and the system thus made unreliable.

The ISOLATION property of ACID is often relaxed (i.e., partly respected) because of the speed decrease this type of concurrency management implies. In effect the only strict way to respect the isolation property is to use a serial model where no two transactions can occur on the same data at the same time and where the result is predictable (i.e. transaction B will happen after transaction A in every single possible case).

In reality, many alternatives are used due to speed concerns, but none of them guarantee the same reliability.

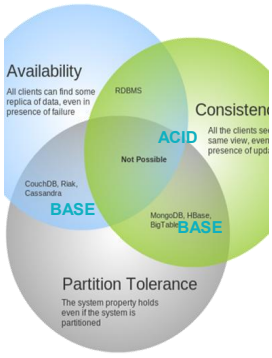
Durability

Durability means that once a transaction has been committed, it will remain so. In other words, every committed transaction is protected against power loss/crash/errors and cannot be lost by the system and can thus be guaranteed to be completed.

In a relational database, for instance, once a group of SQL statements execute, the results need to be stored permanently. If the database crashes right after a group of SQL statements execute, it should be possible to restore the database state to the point after the last transaction committed.

DMA DATA MANAGEMENT ADVISORS

BaSE



CAP computer science theorem quantifies the trade-offs.

Eric Brewer's CAP theorem:
If you want consistency, availability, and partition tolerance, you have to settle for two out of three.

For a distributed system, ***partition tolerance*** means the system will continue to work unless there is a total network failure. A few nodes can fail and the system keeps going.

BASE is an alternative to **ACID**:

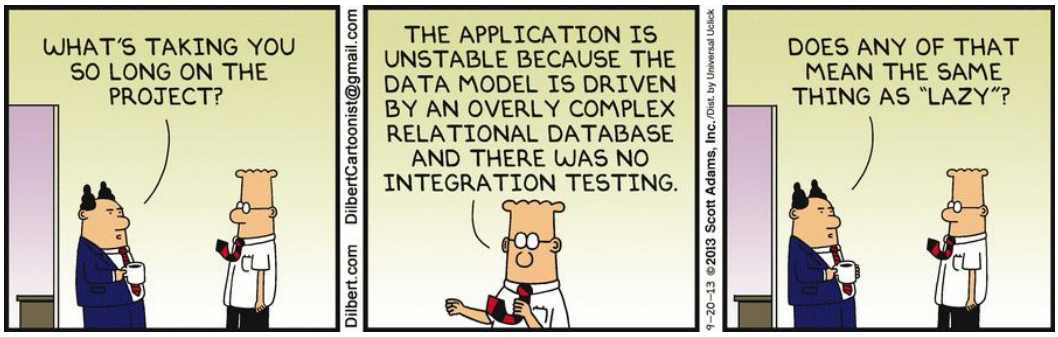
- **Ba**sic Availability
- **S**oft-state
- **E**ventual consistency

Rather than requiring consistency **after every transaction**, it is enough for the database to **eventually be in a consistent state**. *(Accounting systems do this all the time … called "closing out the books.")*

It's OK to use stale data, and it's OK to give approximate answers.

DMA DATA MANAGEMENT ADVISORS


'BOUT THAT BASS, NO TREBLE


WHAT'S TAKING YOU SO LONG ON THE PROJECT?
THE APPLICATION IS UNSTABLE BECAUSE THE DATA MODEL IS DRIVEN BY AN OVERLY COMPLEX RELATIONAL DATABASE AND THERE WAS NO INTEGRATION TESTING.
DOES ANY OF THAT MEAN THE SAME THING AS "LAZY"?
Dilbert.com DilbertCartoonist@gmail.com
9-20-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick

Ref	Question	A	B	C	D	E
DD1	What are relationship labels?	The verb phrases describing the business rules in each direction between two entities.	The nullability setting on a foreign key.	A foreign key that has been role-named.	A relationship without cardinality.	A non-identifying relationship.
DD2	Which of these statements has the most meaningful relationship label?	An order line contains orders	An order is associated with order lines.	An order is related to order lines.	An order is connected with order lines.	An order is composed of order lines.
DD3	All of the following are TRUE statements on relationship types except:	A many-to-many relationship says that an instance of each entity may be associated with many instances of the other entity, and vice versa.	A one-to-one relationship says that a parent entity may have one and only one child entity.	A one-to-many relationship says that a parent entity may have one or more child entities.	A one-to-many relationship says that a child entity may have one or more parent entities.	A recursive relationship relates instances of an entity to other instances of the same entity.
DD4	In the BASE vs ACID model for Transaction Processing, E is best described which of these statements?	End to End data consistency	Eventual Availability of Data as described by the CAP theorem	Eventual Data Consistency	Extra Validation	Business Availability of Secure data Elements
DD5	A bank applies the business rule that each Customer may own one or many Accounts and each Account must be owned by one or many Customers. Which relationship type would be most appropriate?	many-to-many.	one-to-many.	many-to-one.	one-to-one.	recursive.
DD6	An employee may work for one other employee and may manage one or more employees. There is an indeterminate number of levels in this management hierarchy. What type of relationship would work best?	recursive.	one-to-one.	subtyping.	identifying.	non-identifying.
DD7	Which of the following business rules should NOT appear on a logical data model?	Each Person can work for zero to many Companies	Customer Last Name requires a non-unique index to improve retrieval performance.	Each Company must employ one or many Persons.	Each Order can contain one or many Order Lines.	Each Policy must belong to one Policy Owner.
DD8	All of the following are properties of a logical data model except:	contains relationship cardinality	technology-independent	contains attributes.	contains primary keys.	technology-dependent
DD9	According to the DMBoK, the deliverables for a data modelling process do not have to include:	the steps in the business process that use the data.	one or more diagrams.	definitions for entities, attributes and relationships.	issues and outstanding questions regarding data usage.	lineage describing where the data came from.

The deliverables of the data modeling process include:

- **Diagram**: A data model contains one or more diagrams. The diagram is the visual that captures the requirements in a precise form. It depicts a level of detail (e.g., conceptual, logical, or physical), a scheme (relational, dimensional, object-oriented, fact-based, time-based, or NoSQL), and a notation within that scheme (e.g., information engineering, unified modeling language, object-role modeling).
- **Definitions**: Definitions for entities, attributes, and relationships are essential to maintaining the precision on a data model.
- **Issues and outstanding questions**: Frequently the data modeling process raises issues and questions that may not be addressed during the data modeling phase. In addition, often the people or groups responsible for resolving these issues or answering these questions reside outside of the group building the data model. Therefore, often a document is delivered that contains the current set of issues and outstanding questions. An example of an outstanding issue for the student model might be, "If a **Student** leaves and then returns, are they assigned a different **Student Number** or do they keep their original **Student Number**?"
- **Lineage**: For physical and sometimes logical data models, it is important to know the data lineage, that is, where the data comes from. Often lineage takes the form of a source/target mapping, where one can capture the source system attributes and how they populate the target system attributes. Lineage can also trace the data modeling components from conceptual to logical to physical within the same modeling effort. There are two reasons why lineage is important to capture during the data modeling. First, the data modeler will obtain a very strong understanding of the data requirements and therefore is in the best position to determine the source attributes. Second, determining the source attributes can be an effective tool to validate the accuracy of the model and the mapping (i.e., a reality check).

Page 149 / 152 in printed version
Page 156-157 of DM-BOK pdf
Section 2.1 (Activities / Plan for Data Modelling)

AFTER QUIZ 6

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)

Maximum possible score = 43

60% (CDMP Associate) = 26

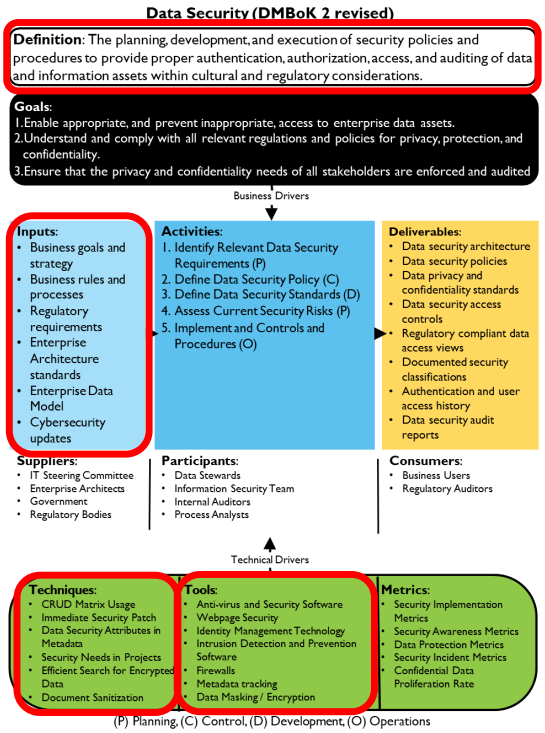
70% (CDMP Practitioner) = 31

80% (CDMP Master) = 35

Data Security Management

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA
DATA MANAGEMENT ADVISORS



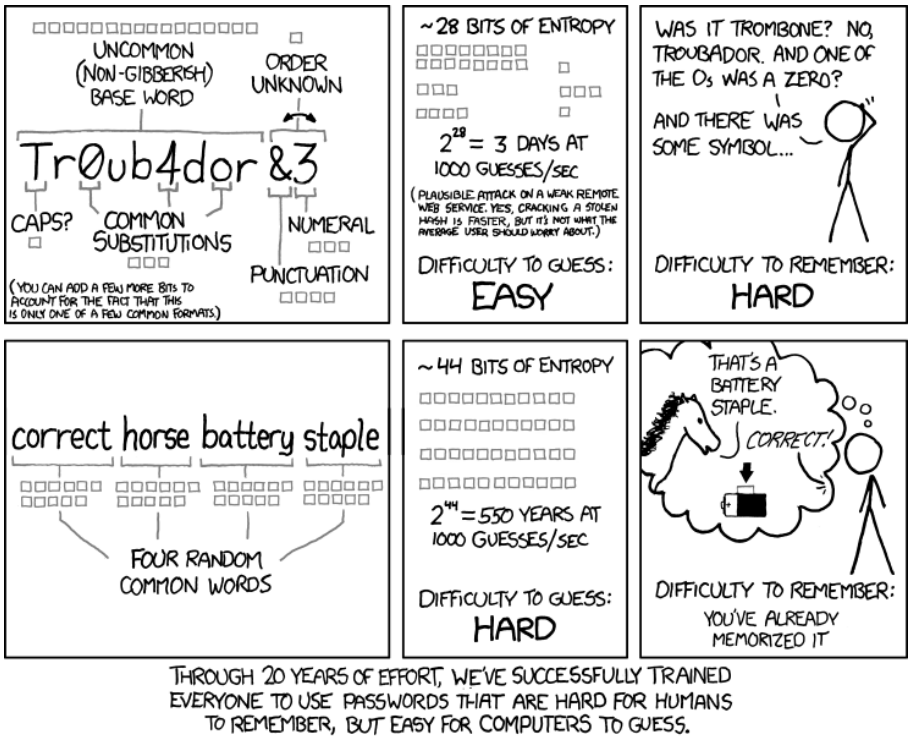
Key Points

- Data Security requirements address “The Four ‘A’s”
  - Authentication Authorization Access Audit
- Data Security is mandated by a range of regulatory provisions across many jurisdictions. It is also basic common sense.
- Breaches of information security can be significantly damaging to the reputation of the organisation, over and above financial or technical damage.
- DMBOK provides a general overview. Prudent Information Managers will refer to relevant ISO standards (ISO27001 / 27701 / 29100) for relevant guidance.

Data Security Guiding Principles

1. Be a responsible trustee of data about all parties
2. Understand and comply with all relevant regulations and guidelines
3. Use CRUD matrices to help map data access needs
4. Ensure Data Security Policy is reviewed and approved by the data governance council (aka DGSC)
5. Identify detailed application security requirements on projects
6. Classify all enterprise data and information products for confidentiality (and other classifications)
7. Set passwords following a set of password complexity guidelines
8. Create security role groups
9. Formally request, track and approve all user and group authorisations
10. Centrally manage user identity data and group membership data
11. Use views or partitions to restrict access to sensitive columns or specific rows
12. Strictly limit and consider every use of shared or service user accounts
13. Monitor data access activity to understand trends

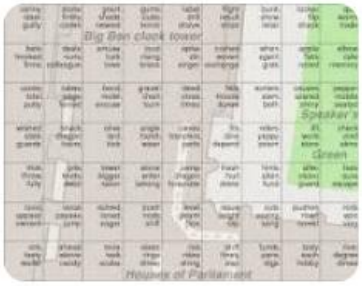
Password Complexity?



https://www.thehindubusinessline.com/info-tech/cyber-security-log4j-vulnerability-issue-explained/article38061525.ece

https://what3words.com/

Wordlists. What3words divides the world into a grid of **57 trillion** 3-by-3-metre (10 ft × 10 ft) squares, each of which has a three-word address. The company says they do their best to remove homophones and spelling variations; however, at least 32 pairs of English near-homophones still remain.



Founders: Chris Sheldrick; Jack Waley-Cohen; ...

Net income: : -£43.3m (2021);

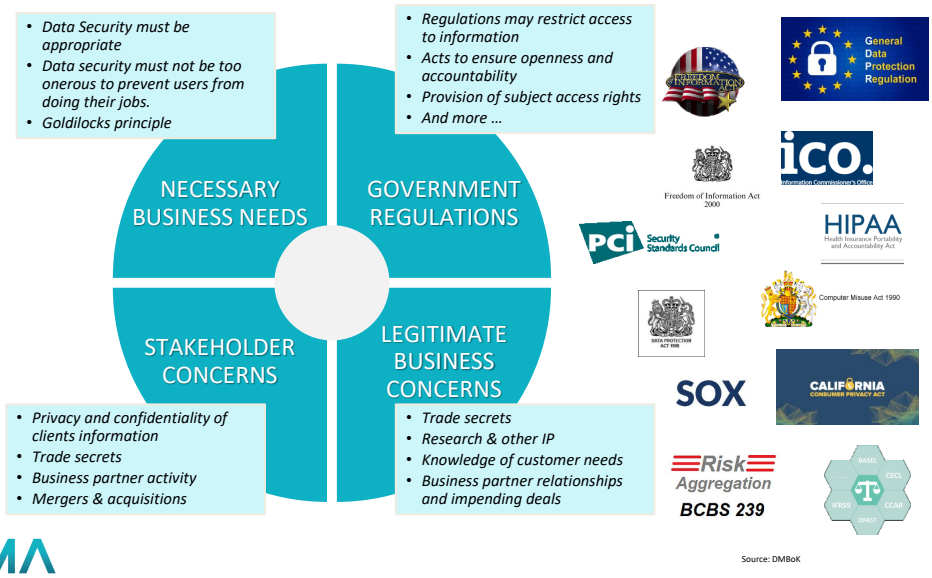
Operating income: : -£38.4m (2021);

https://en.wikipedia.org › wiki › What3words

What3words - Wikipedia

https://what3words.com/fishery.gala.season

Sources of Data Security Requirements



Regulatory and Other Drivers

Security controls are mandated by legislation such as:

- PIPEDA
- PCI-DSS
- EU General Data Protection Regulation (Articles 32-34) GDPR
- HIPAA The Health Insurance Portability and Accountability Act 1996
- Sarbanes-Oxley Act (preventing unauthorised modification of financial transactions)
- BASEL II and Solvency II protecting lineage of data feeding risk models
- CCPA The California Consumer Privacy Act is a state statute intended to enhance privacy rights and consumer protection for residents of California.
- CPRA The California Privacy Rights Act (aka CCPA 2.0)
- BCBS 239 Basel Committee standard number 239 on Banking Supervision. "Principles for effective risk data aggregation and risk reporting“ (RDARR).

Increased consumer and media awareness of data security & the “data foot print” means it must be an active consideration for information managers in Business and IT

Regulation Families

Personal Identification Information (PII):

Also known as Personally Private Information (PPI), includes any information that can personally identify the individual (individually or as a set), such as name, address, phone numbers, schedule, government ID number, account numbers, age, race, religion, ethnicity, birthday, family members' names or friends' names, employment information (HR data), and in many cases, remuneration. Highly similar protective actions will satisfy the EU Privacy Directives, Canadian Privacy law (PIPEDA), PIP Act 2003 in Japan, PCI standards, US FTC requirements, GLB, FTC standards, and most Security Breach of Information Acts.

Financially Sensitive Data:

All financial information, including what may be termed 'shareholder' or 'insider' data, including all current financial information that has not yet been reported publicly. It also includes any future business plans not made public, planned mergers, acquisitions, or spin-offs, non-public reports of significant company problems, unexpected changes in senior management, comprehensive sales, orders, and billing data. All of these can be captured within this one category and protected by the same policies. In the US, this is covered under Insider Trading Laws, SOX (Sarbanes-Oxley Act), or GLBA (Gramm-Leach-Bliley/Financial Services Modernization Act). Note: the Sarbanes-Oxley Act restricts and manages who can change financial data, thus assuring data integrity, while Insider Trading laws affect all those who can see financial data.

Medically Sensitive Data/Personal Health Information (PHI):

All information regarding a person's health or medical treatments. In the US, this is covered by HIPAA (Health Information Portability and Accountability Act). Other nations also have restrictive laws regarding the protection of personal and medical information. As these are evolving, ensure Corporate Counsel is aware of the need to follow legal requirements in a nation in which the organization does business or has customers.

Educational Records:

All information regarding a person's education. In the US, this is covered by FERPA (Family Educational Rights and Privacy Act).

DMA
DATA MANAGEMENT ADVISORS

.. and more

The California Privacy Rights Act (CPRA) – Effective January 1, 2023

The CPRA applies to for-profit businesses that do business in California and meet any of the following:

- Have a gross annual revenue of over $25 million;
- Buy, receive, or sell the personal data of 100,000 or more California residents or households; or
- Derive 50% or more of their annual revenue from selling or sharing California residents' personal data.

Virginia Consumer Data Protection Act (CDPA) – Effective January 1, 2023

The CDPA applies to businesses in Virginia, or businesses that produce products or services that are targeted to residents of Virginia, and that:

- During a calendar year, control or process the personal data of at least 100,000 Virginia residents, or
- Control or process personal data of at least 25,000 Virginia residents and derive over 50% of gross revenue from the sale of personal data.

Colorado Privacy Act (CPA) – Effective July 1, 2023

The CPA applies to organizations that conduct business in Colorado or produce or deliver commercial products or services targeted to residents of Colorado and satisfy one of the following thresholds:

- Control or process the personal data of 100,000 Colorado residents or more during a calendar year, or
- Derive revenue or receive a discount on the price of goods or services from the sale of personal data, and process or control the personal data of 25,000 Colorado residents or more.

Connecticut Act Concerning Personal Data Privacy and Online Monitoring (CTPDA) – Effective July 1, 2023

The CTPDA applies to any business that conducts business in the state, or produces a product or service targeted to residents of the state, and meets one of the following thresholds:

- During a calendar year, controls or processes personal data of 100,000 or more Connecticut residents, or
- Derives over 25% of gross revenue from the sale of personal data and controls or processes personal data of 25,000 or more Connecticut residents.

Utah Consumer Privacy Act (UCPA) – Effective December 31, 2023

The UCPA applies to any business that conducts business in the state, or produces a product or service targeted to residents of the state, has annual revenue of $25,000,000 or more, and meets one of the following thresholds:

- During a calendar year, controls or processes personal data of 100,000 or more Utah residents, or
- Derives over 50% of the gross revenue from the sale of personal data and controls or processes personal data of 25,000 or more Utah residents.

DMA
DATA MANAGEMENT ADVISORS


A4
KT69 OPH

A4


Supreme
Excellent for your needs
A4
80 GSM
500 sheets

A4

AUTHENTICATION

Validate users are who they say they are

AUTHORISATION

Identify the right individuals and grant them the right privileges to specific, appropriate views of data

ACCESS

Enable individuals and their privileges in a timely manner

AUDIT

Review security actions and user activity (to ensure compliance with regulations and conformance with policy and standards)

DMA
DATA MANAGEMENT ADVISORS

Audit … Monitor User Authentication and Access Behaviour

- Monitoring authentication and access behaviour is important for:
  - Compliance auditing
  - Alerting security administrators to problems.
- Monitoring helps detect unusual or suspicious transactions.

- **Active**, real-time monitoring is appropriate for systems containing confidential data such as salary information.
- **Passive** monitoring takes snapshots of the system state at regular intervals and compares levels of activity against benchmarks.

Monitoring can be:

- Application specific.
- Implemented for certain users or role-groups.
- Implemented for certain privileges.
- Used to validate data integrity.
- Used to validate configuration.

DMA
DATA MANAGEMENT ADVISORS

Defining Standards

- There is no one prescribed way of implementing information security controls in an organisation
- ISO standards and other standards do exist that set out critical requirements for Information Security standards in an organisation:
  - ISO/IEC 29100: Information technology - security techniques - privacy framework Standard. For implementing a privacy framework for use when processing personally identifiable information (PII).
  - ISO/IEC 27001: An international standard on how to manage information security.
  - ISO/IEC 27701: An extension of ISO 27001 to provide the data privacy and information security standards required by GDPR. Implementing ISO 27701 will create a Privacy Information Management System (PIMS)
  - PCI-DSS
- Standards can, and do, influence:
  - Access control
  - Use of mobile devices (e.g. BYOD policy) and associated perimeter security
  - Records and device disposal processes and procedures

DMA DATA MANAGEMENT ADVISORS

https://www.gov.uk/government/publications/cyber-essentials-scheme-overview

CIA

CONFIDENTIALITY

Preventing the disclosure of information to unauthorised individuals or systems.

INTEGRITY

Preventing the undetectable modification of information.

AVAILABILITY

Ensuring that information is available where and when it is needed.

DMA DATA MANAGEMENT ADVISORS

3 categories of controls

Controls are approaches, procedures or other strategies used to address one or more of the information security risks. There are three types (DAMA-I):

ADMINISTRATIVE CONTROLS	*"Procedural security"* *Administrative: Training, Induction, Principles, policies, procedures, standards and guidelines*
LOGICAL CONTROLS	*"Software security"* *Logical: Passwords, firewalls, network intrusion detection systems, encryption, access controls lists*
PHYSICAL CONTROLS	*"Workplace security"* *Physical: Doors, locks, CCTV, guards, sprinkler system*

DMA
DATA MANAGEMENT ADVISORS

Risk Assessment:

THREAT
An aspect that might be environmental or manmade that has the potential to compromise the confidentiality, integrity or availability of an information asset

VULNERABILITY
A weakness that could be exploited to compromise the confidentiality, integrity or availability of an information asset

MITIGATION
The activities that can be undertaken, planned or considered to reduce or eliminate the impact of a loss of confidentiality, integrity or availability of an information asset.
Auditors want to see all rows in the security risk assessment matrix completed (even if mitigation is unfeasible) in which case "accept" the risk" is included.

PROBABILITY
the likelihood that a threat will exploit a vulnerability to compromise the confidentiality, integrity or availability of an information asset

IMPACT
A loss of confidentiality, integrity or availability which may result in more significant losses to competitive advantage, revenue, life, property or reputation

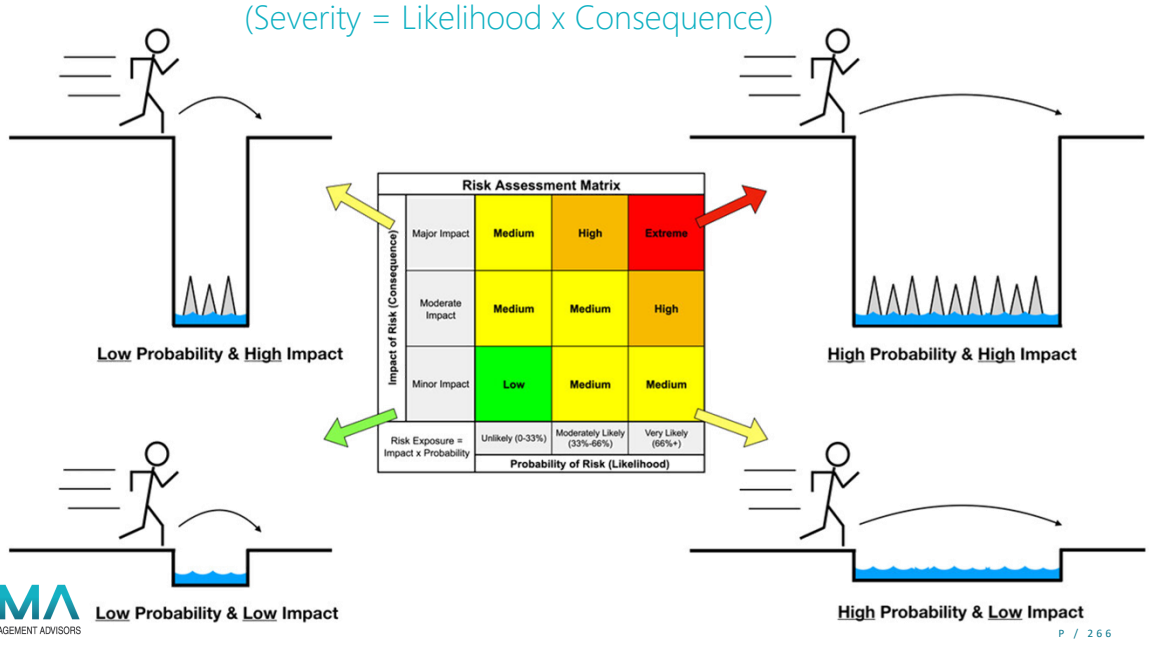
ISO 27001 suggests four ways to treat risks:

1. 'Terminate' the risk by eliminating it entirely,
2. 'treat' the risk by applying security controls,
3. 'transfer' the risk to a third party, or
4. 'tolerate' the risk.

DMA
DATA MANAGEMENT ADVISORS

Assessment of Risk Exposure = Risk Probability x Impact

(Severity = Likelihood x Consequence)



DMA DATA MANAGEMENT ADVISORS

Ref: Mark Warner | TheProjectManagementBlueprint.Com

Risk Classifications

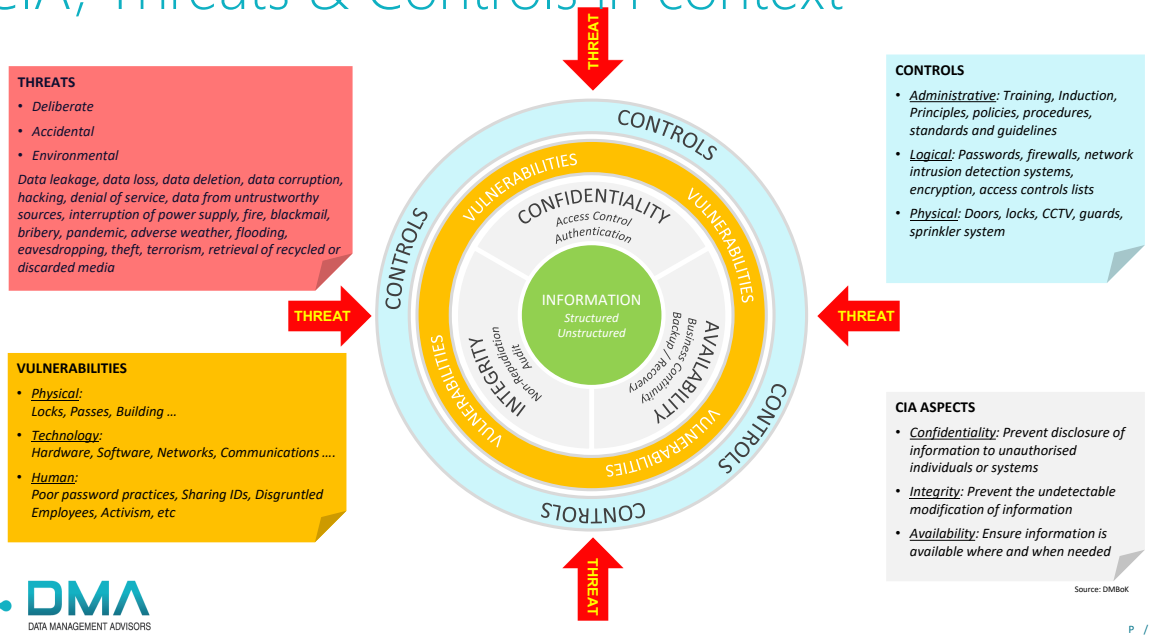
Critical Risk Data (CRD):
Personal information aggressively sought for unauthorized use by both internal and external parties due to its high direct financial value. Compromise of CRD would not only harm individuals, but would result in financial harm to the company from significant penalties, costs to retain customers and employees, as well as harm to brand and reputation.

High Risk Data (HRD):
HRD is actively sought for unauthorized use due to its potential direct financial value. HRD provides the company with a competitive edge. If compromised, it could expose the company to financial harm through loss of opportunity. Loss of HRD can cause mistrust leading to the loss of business and may result in legal exposure, regulatory fines and penalties, as well as damage to brand and reputation.

Moderate Risk Data (MRD):
Company information that has little tangible value to unauthorized parties; however, the unauthorized use of this non-public information would likely have a negative effect on the company.

DMA DATA MANAGEMENT ADVISORS

CIA, Threats & Controls in context



THREATS

- *Deliberate*
- *Accidental*
- *Environmental*

Data leakage, data loss, data deletion, data corruption, hacking, denial of service, data from untrustworthy sources, interruption of power supply, fire, blackmail, bribery, pandemic, adverse weather, flooding, eavesdropping, theft, terrorism, retrieval of recycled or discarded media

CONTROLS

- *Administrative: Training, Induction, Principles, policies, procedures, standards and guidelines*
- *Logical: Passwords, firewalls, network intrusion detection systems, encryption, access controls lists*
- *Physical: Doors, locks, CCTV, guards, sprinkler system*

VULNERABILITIES

- *Physical: Locks, Passes, Building …*
- *Technology: Hardware, Software, Networks, Communications ….*
- *Human: Poor password practices, Sharing IDs, Disgruntled Employees, Activism, etc*

CIA ASPECTS

- *Confidentiality: Prevent disclosure of information to unauthorised individuals or systems*
- *Integrity: Prevent the undetectable modification of information*
- *Availability: Ensure information is available where and when needed*

Source: DMBoK

DMA
DATA MANAGEMENT ADVISORS

IT Security Network Components

Firewall:
A firewall is software and/or hardware that filters network traffic to protect an individual computer or an entire network from unauthorized attempts to access or attack the system. A firewall may scan both incoming and outgoing communications for restricted or regulated information and prevent it from passing without permission (Data Loss Prevention). Some firewalls also restrict access to specific external websites.

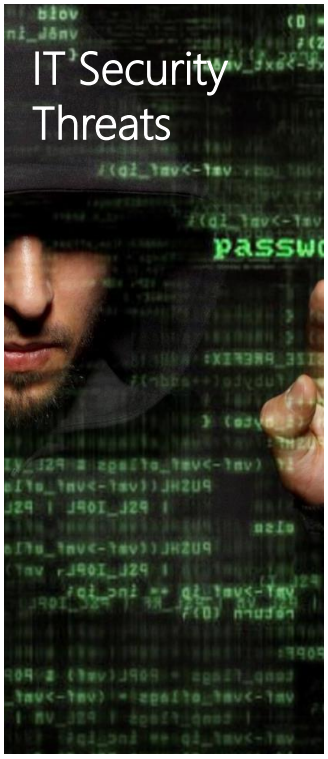
Perimeter:

A *perimeter* is the boundary between an organization's environments and exterior systems. Typically, a firewall will be in place between all internal and external environments.

DMZ:

A de-militarized zone (*DMZ)* is an area on the edge or perimeter of an organization, with a firewall between it and the organization. A DMZ environment will always have a firewall between it and the internet. DMZ environments are used to pass or temporarily store data moving between organizations.

DMA
DATA MANAGEMENT ADVISORS



- **Privilege Escalation**
  Software programs often have bugs that can be exploited. These bugs can be used to gain access to certain resources with higher privileges that can bypass security controls.

- **Virus**
  A virus is a computer program that, like a medical virus, can replicate and infect other computers.

- **Botnets**
  Botnets are created with a Trojan and reside on IRC networks. The bot can launch an IRC client and join chat room in order to spam and launch denial of service attacks.

- **Adware**
  Like spyware, adware observes a user's Internet browsing habits. But the purpose is to be able to better target the display of web advertisements.

- **Worm**
  A worm is a specific type of virus. Unlike a typical virus, its goal isn't to alter system files, but to replicate so many times that it consumes hard disk space or memory.

- **Trojan**
  They masquerade as normal, safe applications, but their mission is to allow a hacker remote access to your computer. In turn, the infected computer can be used as part of a denial-of-service attack and data theft can occur (e.g., keystroke logger).

- **Spyware**
  Like Trojans, spyware can pilfer sensitive information, but are often used as advertising tools as well. The intent is to gather a user's information by monitoring Internet activity and transmitting that to an attacker.

- **Logic bomb**
  They are bits of code added to software that will set off a specific function. Logic bombs are similar to viruses in that they can perform malicious actions like deleting files and corrupting data.

- **Spam**
  is unsolicited junk mail. It comes in the form of an advertisement, and in addition to being a time waster, has he ability to consume precious network bandwidth.

- **Rootkits**
  Rootkits are some of the most difficult to detect. They are activated when your system boots up — before anti-virus software is started. Rootkits allow the installation of files and accounts, or the purposes of intercepting sensitive information.

- **Ransomware**
  Ransomware is a subset of malware in which the data on a victim's computer is locked, typically by encryption, and payment is demanded before the ransomed data is decrypted and access returned to the victim..

- **Zero-day**
  A software vulnerability unknown to those who should be interested in its mitigation (including the vendor of the target software). Until the vulnerability is mitigated, hackers can exploit it to adversely affect programs, data, additional computers or a network. An exploit directed at a zero-day is called a zero-day exploit, or zero-day attack.

DMA
DATA MANAGEMENT ADVISORS

Ref	Question	A	B	C	D	E
DS1	Which of these statements best defines Data Security Management?	The planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets	The implementation and execution of checkpoints, checklists, controls, and technical mechanisms to govern the access to information in an enterprise	The definition of controls, technical standards, frameworks, and audit trail capabilities to identify who has or has had access to information	The planning, implementation, and testing of security technologies, authentication mechanisms, and other controls to prevent access to information	Ensuring that Data is Authenticated, Authorised, Accessed and Audited
DS2	Which of these are characteristics of an effective data security policy?	The procedures defined are benchmarked, supported by technology, framework based, and peer reviewed	The defined procedures are tightly defined, with rigid and effective enforcement sanctions, and alignment with technology capabilities	The policies are specific, measurable, achievable, realistic, and technology aligned	The defined procedures ensure that the right people can use and update data in the right way, and that all inappropriate access and update is restricted	The GDPR and FOI regulations are followed
DS3	Apart from security requirements internal to the organisation, what other strategic goals should a Data Security Management system address?	Compliance with ISO29100 and PCI-DSS	Compliance with ISO27001 and HIPAA	Regulatory requirements for privacy and confidentiality AND Privacy and Confidentiality needs of all stakeholders	Ensuring the organisation doesn't engage in SPAM marketing	Ensuring data breaches are reported to the information commissioner in a timely manner
DS4	The implementation and administration of database security is often the responsibility of ….	The CIO	The Database Administrator	The Database system owner	The Data Governance Council	The System Owner (AKA Custodian)
DS5	What is the role of the Data Governance Council in defining an Information Security policy?	The Data Governance Council should review and approve the high-level Data Security Policy	The Data Governance Council should define the Data Security Policy	The Data Governance Council should implement the Data Security Policy	The Data Governance Council should have no role in Data Security	The Data Governance Council should audit Data Security breaches
DS6	What is the benefit of using role groups to implement data security policies?	It simplifies revoking individual permissions from an individual user	It allows users to be typecast by the administrator	It reduces the amount of effort to assign access rights to users if they inherit rights from their group	It allows for iterative reporting of user access	None, security policies should be set according to individual user needs

AFTER QUIZ 7

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)

Maximum possible score = 49

60% (CDMP Associate) = 30

70% (CDMP Practitioner) = 35

80% (CDMP Master) = 40

Document, Records & Content Management

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

Document and Content Management (DMBoK 2 revised)

Definition: Controlling the capture, storage, access, and use of data and information stored predominantly outside relational databases.

Goals:
1. To comply with legal obligations and customer expectations regarding Records management.
2. To ensure effective and efficient storage, retrieval, and use of Documents and Content.
3. To ensure integration capabilities between structured and unstructured Content.

Business Drivers

Inputs:
- Business strategy
- IT strategy
- Legal retention requirements
- Text file
- Electronic format file
- Printed paper file
- Social media stream

Activities:
1. **Plan for Lifecycle Management (P)**
   1. Plan for Records Management
   2. Develop a content strategy
   3. Create Content Handling Policies
   4. Define Content Information Architecture
2. **Manage the Lifecycle (O)**
   1. Capture Records and Content (O)
   2. Manage Versioning and Control
   3. Backup and Recovery
   4. Manage Retention and Disposal
   5. Audit Document and Records
3. **Publish and Deliver Content (O)**
   1. Provide Access, Search, and Retrieval
   2. Deliver Through Acceptable Channels

Deliverables:
- Content and Records Management Strategy
- Policy and procedure
- Content Repository
- Managed record in many media formats
- Audit trail and log
- E-discovery approaches

Suppliers:
- Legal team
- Business team
- IT team
- External party

Participants:
- Data steward
- Data management professional
- Records management staff
- Content management staff
- Web development staff
- Librarians

Consumers:
- Business user
- IT user
- Government regulatory agency
- Audit team
- External customer

Technical Drivers

Techniques:
- Metadata tagging
- Data markup and exchange format
- Data mapping
- Storyboarding
- Infographics

Tools:
- Office productivity software
- Enterprise content management system
- Controlled vocabulary / meta-data tool
- Knowledge management wiki
- Visual media tool
- Social media
- E-discovery technology

Metrics:
- Compliance audit metric
- Return on investment
- Usage metric
- Record management KPI
- E-discovery KPI
- ECM program metric
- ECM operational metric

(P) Planning, (C) Control, (D) Development, (O) Operations

Main Activities

Document & Records Management

- **Document / Record Management** is the lifecycle management of the designated significant documents of the organization.
- Not all documents are significant as evidence of the organization's business activities and regulatory compliance.
- Records management manages paper and microfiche / film records from their creation or receipt through processing, distribution, organization, and retrieval, to their ultimate disposition.

Content Management

DMA DATA MANAGEMENT ADVISORS

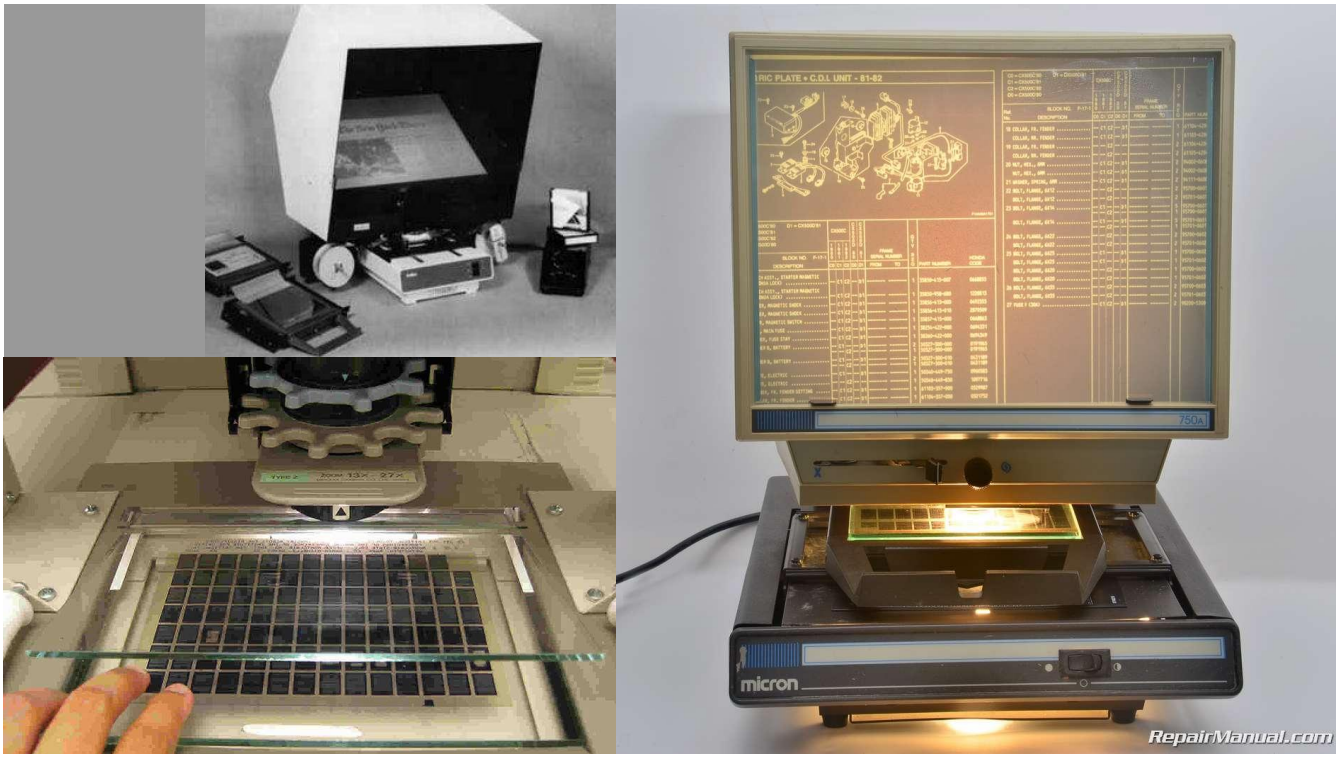
Key Terms

- **Content Management**
  The organisation, categorisation, and structure of data / resources so that they can be stored, published and reused in multiple ways.
- **Document Management**
  The storage, inventory and control of electronic and paper documents.
- **Records Management**
  A subset of Document management. Includes the full lifecycle: from record creation or receipt through processing, distribution, organization, and retrieval, to disposition.

  Records can be physical (e.g., documents, memos, contracts, reports or microfiche); electronic (e.g., email content, attachments, and instant messaging); content on a website; documents on all types of media and hardware; and data captured in databases of all kinds.

  Hybrid records combine formats, e.g. aperture cards (paper record with a microfiche window imbedded with details or supporting material).
- **Significant Records / Documents**
  Significant as evidence of the organization's business activities and regulatory compliance.
- **Vital Record**
  A type of record required to resume an organization's operations the event of a disaster.
- **Taxonomy**
  The science or technique of classification.
- **Ontology**
  A type of model that represents a set of concepts and their relationships within a domain.

DMA DATA MANAGEMENT ADVISORS



Appliance Category	Class 1		Class 2	
	Visual inspection	PAT test	Visual inspection	PAT test
Fixed	2 yearly	Not required	2 yearly	Not required
Stationary	2 yearly	4 yearly	2 yearly	Not required
IT	2 yearly	4 yearly	2 yearly	Not required
Moveable	Annually	4 yearly	2 yearly	Not required
Portable	Annually	4 yearly	2 yearly	Not required
Cables & Chargers	Annually	4 yearly	2 yearly	Not required
Handheld	6 monthly	Annually	6 monthly	Not required

DMA
DATA MANAGEMENT ADVISORS

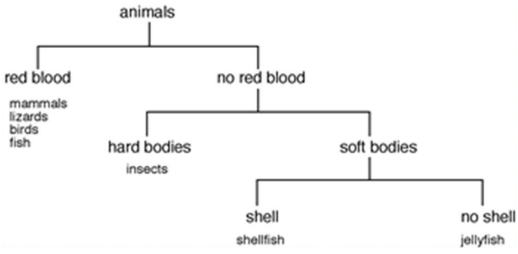
TAXONOMIES

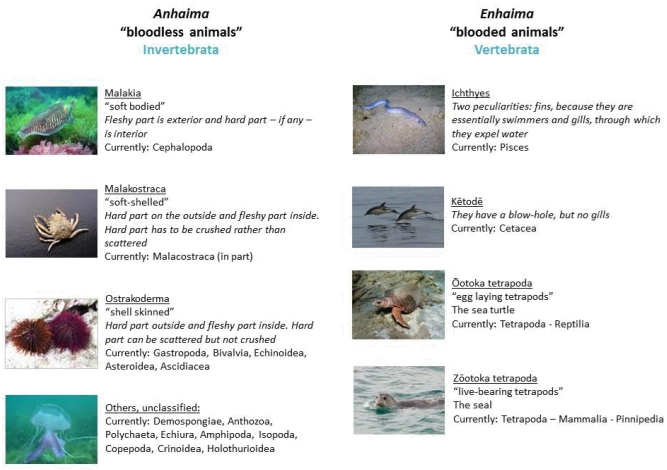
Grouped into five types:

1. **Flat Taxonomy** – no relationship among the controlled set of categories (example: list of countries).
2. **Hierarchical Taxonomy** – tree structure; for example geography, from continent down to address.
3. **A Polyhierarchy** is a tree-like structure with more than one node relation rule. Child nodes may have multiple parents. Those parents may also share grandparents.
4. **Facet Taxonomy** – looks like a star where each node is associated with the centre node. Facets are attributes of the object in the centre; for example meta-data where each attribute (creator, title, keywords etc.) is a facet of a content object.
5. **Network Taxonomy** – uses both uses both hierarchical and facet structures, for example a thesaurus, or recommender engine (if you liked that, you may also like this…).

DMA
DATA MANAGEMENT ADVISORS

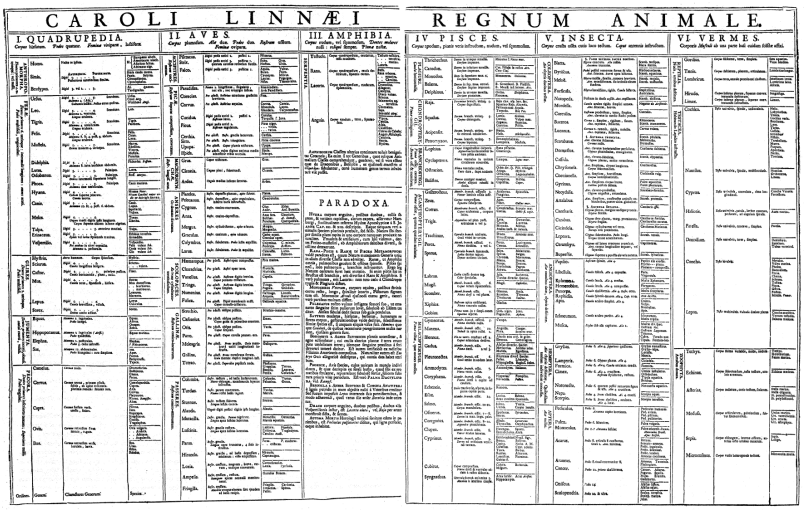
Aristotle's classification system (384–322 BC)





DMA
DATA MANAGEMENT ADVISORS

Systema Naturae Carl Linnaeus, 1735


CAROLI LINNÆI

I QUADRUPEDIA.

II AVES

III AMPHIBIA.

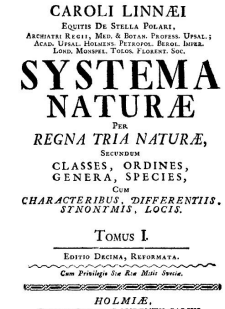
REGNUM ANIMALE.

IV PISCES.

V INSECTA.

VI VERMES.

PARADOXA


CAROLI LINNÆI
EQUITIS DE STELLA POLARI,
ARCHIATRI REGII, MED. & BOTAN. PROFESS. UPSAL.;
ACAD. UPSAL. HOLMENS. PETROPOL. BEROL. IMPER.
LOND. MANSFIEL. TUSCIN. FLORENT. SOC.

SYSTEMA NATURÆ

PER
REGNA TRIA NATURÆ,
SECUNDUM
CLASSES, ORDINES,
GENERA, SPECIES,
CUM
CHARACTERIBUS, DIFFERENTIIS,
SYNONYMIS, LOCIS.

TOMUS I.

EDITIO DECIMA, REFORMATA.

Cum Privilegio Sa. Re. Mtis. Suecie.

HOLMIÆ,
IMPENSIS DIRECT. LAURENTII SALVII,
1758.

DMA
DATA MANAGEMENT ADVISORS

1. ***An ontology is a type of taxonomy*** that represents a set of concepts and their relationships within a domain.
2. Ontologies provide the primary knowledge representation in the Semantic Web and are used in the exchange of information between Semantic Web applications.
3. Ontologies describe classes (concepts), individuals (instances), attributes, relations, and events.
4. Two main differences between Taxonomy & Ontology
   a) A taxonomy provides data content classifications for a given concept area. A data model specifically calls out the entity an attribute belongs to. In an ontology, entity, attribute, and content concepts can be completely mixed. Differences are identified through Metadata or other relationships.
   b) In a taxonomy or data model, what is defined is what is known – and nothing else. This is referred to as a **closed-world assumption**. In an ontology, ***possible*** relationships are inferred based on the nature of existing relationships, so something that is not explicitly declared can be true. This is referred to as the **open-world assumption**.

Example Ontologies

FIBO The Financial Industry Business Ontology (FIBO) is a business conceptual model showing how all financial instruments, business entities and processes work in the financial industry.

CHEMINF The chemical information ontology (cheminf) – contains information entities about chemical entities & includes terms for the descriptors commonly used in cheminformatics software applications and the algorithms which generate them..

FMA - Foundational Model of Anatomy A domain ontology that represents a coherent body of explicit declarative knowledge about human anatomy. Its ontological framework can be applied and extended to all other species.

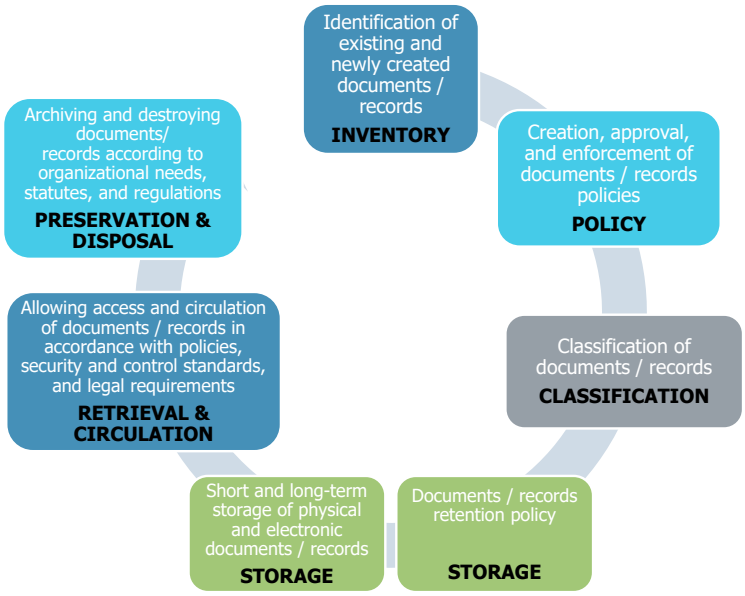
GO - Gene Ontology A collaborative effort to address the need for consistent descriptions of gene products across databases. The GO represents information about biological processes, cellular components and molecular functions.

HPO - Human Phenotype Ontology An ontology that represents phenotypic abnormalities encountered in human disease.

IDO - Infectious Disease Ontology A set of interoperable ontologies covering the infectious disease domain. At the core of the set is a general Infectious Disease Ontology (IDO-Core) of entities relevant to both biomedical and clinical aspects of most infectious diseases. Sub-domain specific extensions of IDO-Core provide coverage of entities relevant to specific pathogens or diseases.

DMA
DATA MANAGEMENT ADVISORS

Document / Record Management Lifecycle



DMA
DATA MANAGEMENT ADVISORS

Document Control Schemes

- Based on the criticality of the data and the perceived harm that would occur if data were corrupted or otherwise unavailable.
- ANSI Standard 859 (2008) describes three levels of control:
  - **Custody control** is the least formal, merely requiring safe storage and a means of retrieval.
  - **Revision control** is more formal, notifying stakeholders and incrementing versions when a change is required. **Requires Administrative Metadata**
  - **Formal control** requires formal change initiation, thorough change evaluation for impact, decision by a change authority, and full status accounting of implementation and validation to stakeholders. **Requires Administrative Metadata**

Document and Records Management Overview

- Lifecycle management of significant organizational documents
- Includes planning, capturing, versioning, retention, and disposal
- Ensures compliance with legal and regulatory requirements
- Supports access, search, retrieval, and audit processes
- Involves collaboration among legal, IT, and business teams

DMA
DATA MANAGEMENT ADVISORS

Key Principles and Practices in Document Management

- Document control schemes include custody, revision, and formal control
- Classification and taxonomy define document organization and retrieval
- Retention and disposal policies maintain compliance and reduce risk
- Vital records are critical for business continuity post-disaster
- Content management enables reuse and structured publishing

DMA
DATA MANAGEMENT ADVISORS

Effective Document and Records Management

- Clear policies and procedures are essential for retention and disposal.
- Retention policies define maintenance timeframes for valuable documents.
- Inactive documents may be transferred to secondary or off-site storage.
- Oversight ensures privacy, data protection, and identity theft prevention.
- Software versions impact access and readability of electronic records.
- Non-value-added information should be removed to save space and cost.
- Risks exist when retaining records beyond legally required timeframes.
- Challenges include inadequate policies and lack of buy-in for records management.

Many organizations do not prioritize the removal of non-value-added information because:

- Policies are not adequate
- One person's non-valued-added information is another's valued information
- Inability to foresee future possible needs for current non-value-added physical and / or electronic records
- There is no buy-in for Records Management
- Inability to decide which records to delete
- Perceived cost of deciding and removing physical and electronic records
- Electronic space is cheap. Buying more space when required is easier than archiving and removal processes

DMA
DATA MANAGEMENT ADVISORS

Generally Accepted Recordkeeping Principles GARP*

The Principle of Accountability - A senior executive (or a person of comparable authority) shall oversee the information management to appropriate individuals.

The Principle of Transparency - An organization's business processes and activities, including its **governance** program, shall be documented in an open and verifiable manner, and that documentation shall be available to all personnel and appropriate, interested parties.

The Principle of Integrity - A **governance** program shall be constructed so the information assets generated by or managed for the organization have a reasonable guarantee of authenticity and reliability.

The Principle of Protection - A **governance** program shall be constructed to ensure an appropriate level of protection to information assets that are private, confidential, privileged, secret, classified, essential to business continuity, or that otherwise require protection.

The Principle of Compliance - A **governance** program shall be constructed to comply with applicable laws, other binding authorities, and the organization's policies.

The Principle of Availability - An organization shall maintain its information assets in a manner that ensures their timely, efficient, and accurate retrieval.

The Principle of Retention - An organization shall maintain its information assets for an appropriate time, taking into account its legal, regulatory, fiscal, operational, and historical requirements.

The Principle of Disposition - An organization shall provide secure and appropriate disposition for information assets no longer required to be maintained, in compliance with applicable laws and the organization's policies.

DMA
DATA MANAGEMENT ADVISORS

** ARMA International*

Main Activities

Document & Records Management

European Commission Article 29 Data Protection Working Party criteria to evaluate anonymization methods.

Anonymity is protected when it is only possible to analyse sizeable "clusters" of individuals who cannot be distinguished from one another based on their attributes

Content Management

- **Content management** is the organization, categorization, and structure of data / resources to be stored, published, and reused in multiple ways.
- **Content** includes data / information, that exists in many forms and in multiple stages of completion within its lifecycle. Content may be found on electronic, paper or other media.
- **The lifecycle of content** can be active, with daily changes through controlled processes for creation, modification, and collaboration of content before dissemination.

DMA DATA MANAGEMENT ADVISORS

Content Delivery Methods

Push: In a push delivery system, users choose the type of content delivered to them on a predetermined schedule. Syndication involves one party creating the content published in many places. Really Simple Syndication (RSS) is an example of a push content delivery mechanism. It distributes content (i.e., a feed) to syndicate news and other web content upon request.

Pull: In a pull delivery system, users pull the content through the Internet. An example of a pull system is when shoppers visit online retail stores.

Interactive: Interactive content delivery methods, such as third-party electronic point of sale (EPOS) apps or customer facing websites (e.g., for enrolment), need to exchange high volumes of real-time data between enterprise applications. Options for sharing data between applications include Enterprise Application Integration (EAI), Changed Data Capture, Data Integration and EII.

DMA DATA MANAGEMENT ADVISORS

Ref	Question	A	B	C	D	E
DOC1	Which of the following are primary deliverables of proper document and record management?	Data from tracking devices, building sensor data	Relational databases, database logs, paper documents	Local drives of laptops, transcripts of phone calls	Spreadsheets, company library books, sales transactions	Managed records in many media formats, e-discovery records, policies and procedures, contracts and financial documents
DOC2	Non-value-added information is often not removed because	The policies are unclear of what is defined as non-value-added so there is no cost driver, and it takes more effort to dispose than to keep.	We might need the information at a later stage	Data is an asset. It is likely to be recognized as valuable in the future.	Legislation is unclear on what should be kept	It should not be removed. All data is value-added
DOC3	When defining your business continuity plan, which of the following should one consider doing?	Have the contracts in place to acquire new hardware in case of technical problems, define policies	Write a report and discuss with management the required budget	Make sure that the data is retained sufficiently long, check that critical data is encrypted, check access rights	Determine the risk, probability and impact, check document backup frequency	Consider written policies and procedures, impact mitigating measures, required recovery time and acceptable amount of disruption, the criticality of the documents

DMA
DATA MANAGEMENT ADVISORS

AFTER QUIZ 8

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)
8. Document, Records & Content Mgt (3)

Maximum possible score = 52

60% (CDMP Associate) = 32

70% (CDMP Practitioner) = 37

80% (CDMP Master) = 42

Data Architecture & Lifecycle Management

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

Data Architecture (DMBoK 2 revised)

Definition: Identifying the data needs of the enterprise (regardless of structure) and designing and maintaining the master blueprints to meet those needs. Using master blueprints to guide data integration, control data assets, and align data investments with business strategy.

Goals:
1. Identify data storage and processing requirements.
2. Strategically prepare organizations to quickly evolve their products, services, and data to take advantage of business opportunities inherent in emerging technologies
3. Design structures and plans to meet the current and long-term data requirements of the enterprise.

Business Drivers

Inputs:
- Enterprise Architecture
- Business Architecture
- IT Standards and Goals
- Data Strategies

Activities:
1. **Establish Enterprise Data Architecture (P)**
   1. Evaluate Existing Data Architecture Specifications
   2. Develop a Roadmap
2. **Manage Enterprise Requirements within Projects (D)**
3. **Integrate with Enterprise Architecture (O)**

Deliverables:
- Data Architecture Design
- Data Flows
- Data Value Chains
- Enterprise Data Model
- Implementation Roadmap

Suppliers:
- Enterprise Architects
- Data Stewards
- Subject Matter Experts
- Data Analysts

Participants:
- Enterprise Data Architects
- Data Modelers

Consumers:
- Database Administrators
- Software Developers
- Project Managers
- Support Teams

Technical Drivers

Techniques:
- Lifecycle Reviews
- Diagramming Clarity

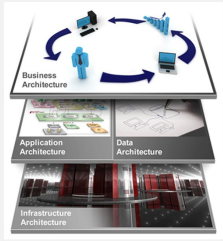
Tools:
- Data modeling tools
- Asset management software
- Graphical design applications

Metrics:
- Architecture standards compliance rates
- Trends in implementation
- Business value metrics

(P) Planning, (C) Control, (D) Development, (O) Operations

Essential Concepts

- Data Architecture appears in all EA Frameworks



- EA Working in practice …
  - Data model tells you the entity / attribute
  - Process model tells you the action
  - System tells you how the application is programmed to implement the required data processes and handle the data
  - Technology tells you how you are delivering the data management experience to people and other systems
  - Confirmed by the business, application, data and the technology architect

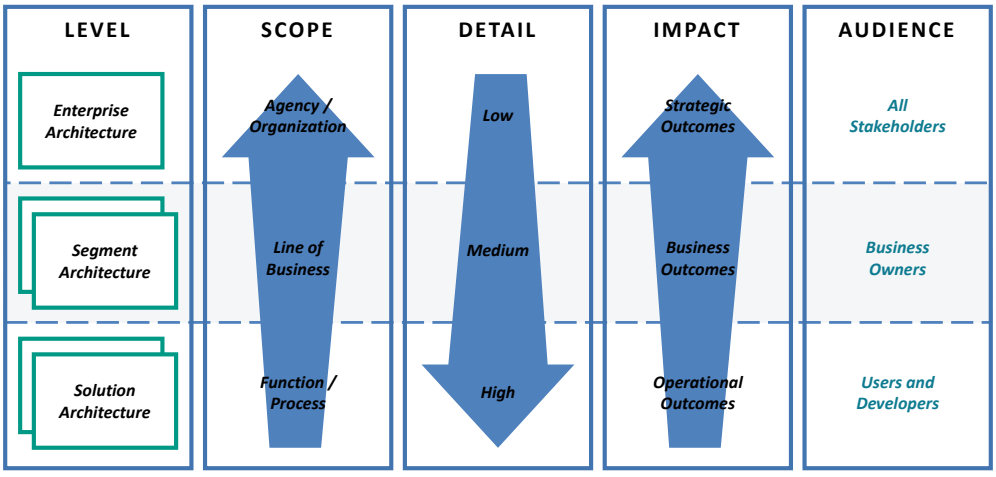
DMA DATA MANAGEMENT ADVISORS

Enterprise Architecture Types & Structures

Enterprise Architecture
Enterprise architecture (EA) is the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving the key requirements, principles and models that describe the enterprise's future state and enable its evolution.

Segment Architecture
Segment architecture is a detailed, formal description of areas within an enterprise, used at the program or portfolio level to organize and align change activity.

Solutions Architecture
Solution architecture is a kind of architecture domain, that aims to address specific problems and requirements, usually through the design of specific information systems or applications.

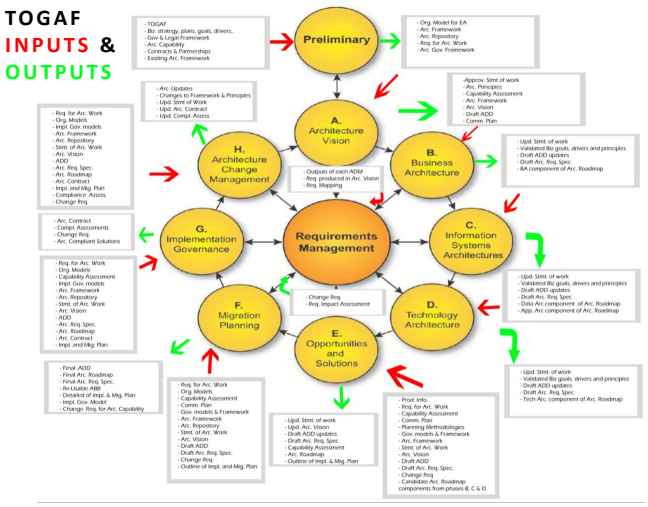


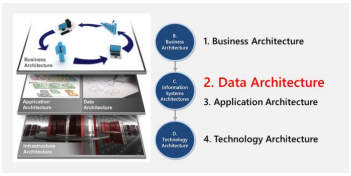
DMA DATA MANAGEMENT ADVISORS

Enterprise Architecture Frameworks

TOGAF

The Open Group Architecture Framework, probably the most widely adopted framework, contains an Architecture Development Method (ADM), content metamodel, and defined artefacts within the business, application, data, and technology domains.



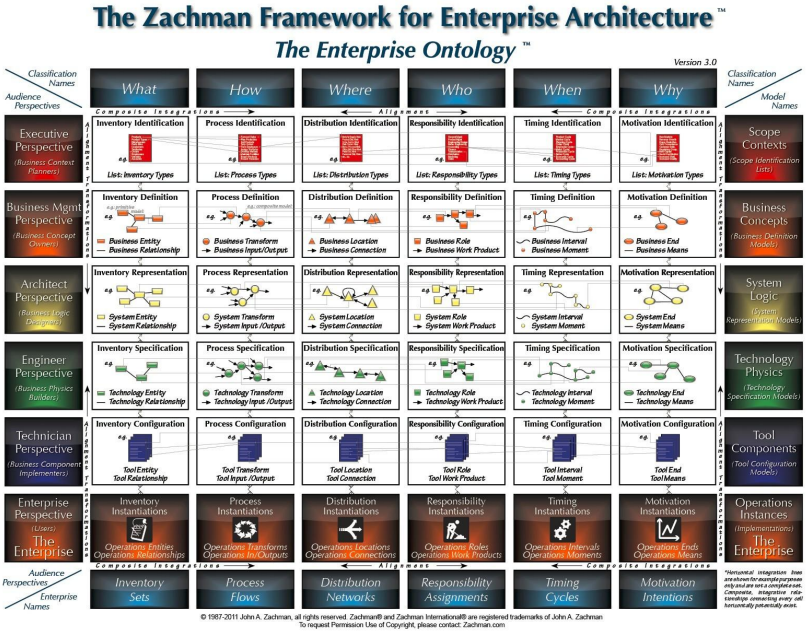


DMA
DATA MANAGEMENT ADVISORS

Enterprise Architecture Frameworks

Zachman

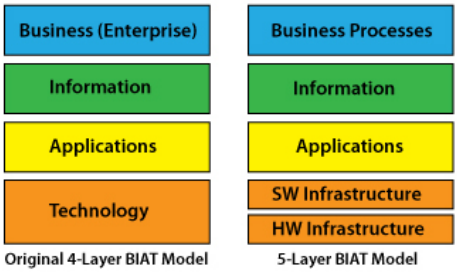
The first enterprise architecture framework defines artefacts in a 6 x 6 matrix, with interrogatives (what, how, where, etc.) as columns and stakeholder perspective as rows (executive, business, architect, etc.). It is an ontology, not a methodology for enterprise architecture.



DMA
DATA MANAGEMENT ADVISORS

Domain	Enterprise Business Architecture	Enterprise Data & Information Architecture	Enterprise Applications Architecture	Enterprise Technology Architecture
Purpose	To identify how an enterprise creates value for customers and other stakeholders	To describe how data should be organized and managed	To describe the structure and functionality of applications in an enterprise	To describe the physical technology needed to enable systems to function and deliver value
Elements	Business models, processes, capabilities, services, events, strategies, vocabulary	Data models, data definitions, data mapping specifications, data flows, structured data APIs	Business systems, software packages, databases	Technical platforms, networks, security, integration tools
Dependencies	Establishes requirements for the other domains	Manages data created and required by business architecture	Acts on specified data according to business requirements	Hosts and executes the application architecture
Roles	Business architects and analysts, business data stewards	Data architects and modelers, business data stewards	Application architects, Solution architects	Infrastructure architects, Solution architects

Enterprise Architecture Domains



DMA
DATA MANAGEMENT ADVISORS

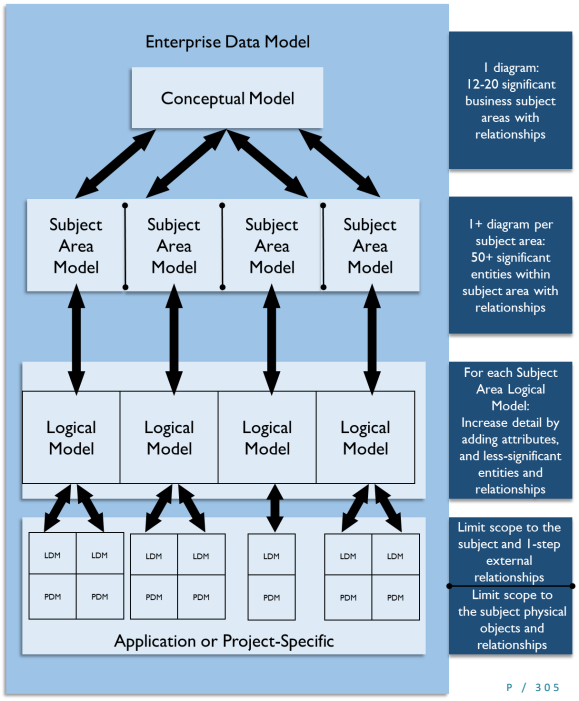
Types of Models in a Typical Data Architecture

Some organizations create an EDM as a stand-alone artifact.

In other organizations, it is composed of data models from different perspectives and at different levels of detail, that consistently describe an organization's understanding of data entities, data attributes, and their relationships across the enterprise.

An EDM includes both universal (Enterprise-wide Conceptual and Logical Models) and application- or project specific data models, along with definitions, specifications, mappings, and business rules.

Adopting an industry standard model (AKA Consensus model) can jumpstart the process of developing an EDM.



DMA
DATA MANAGEMENT ADVISORS

Process and Data ARE related

Create
Read
Update
Delete

Major Entities / Data Subject Areas	Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing
	Business Processes						
Product	C	R	U	U	U		
Product Part	C	R	R	U	U		
Manufacturing Plant	U		C	R	R	U	
Customer	R	C		U	R	U	U
Sales Item	C	C	C	U		U	U
Assembly Structure	U		C		U		
Sales Order		U		R	U	U	U
Production Order			U	C	U	U	U
Individual Product					C	R	U
Shipping						C	
Customer's Invoice		U					C

Process and Data ARE related

Create
Read
Update
Delete

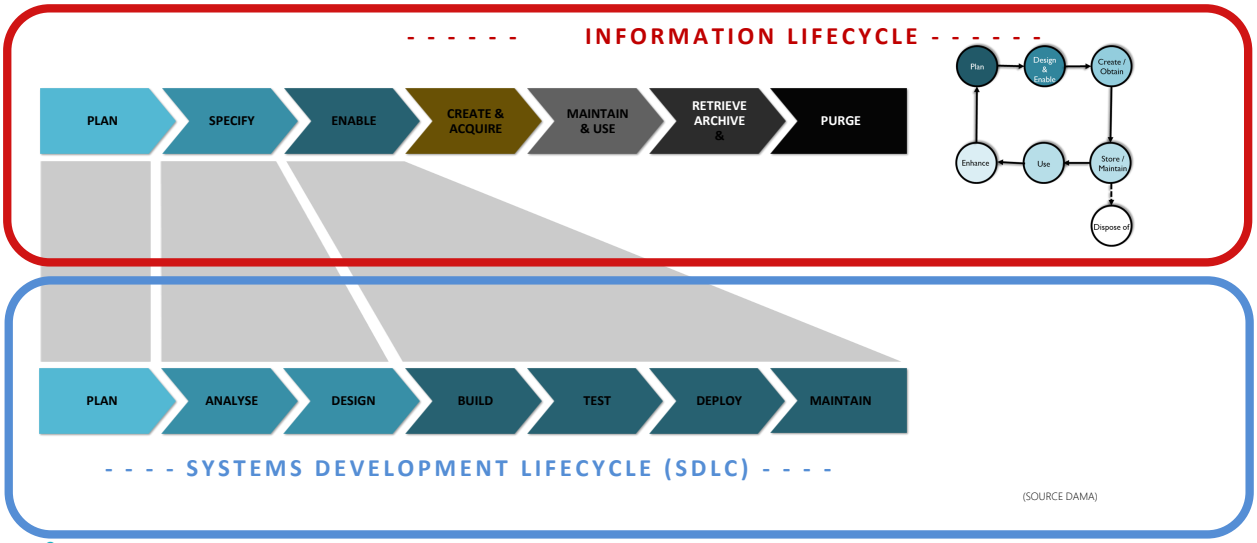
	Major Entities / Data Subject Areas	Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing
		Business Processes						
	Product	C	R	U	U	U		
	Product Part	C	R	R	U	U		
	Manufacturing Plant	U		C	R	R	U	
	Customer	R	C		U	R	U	U
1	Sales Item	C	C	C	U		U	U
	Assembly Structure	U		C		U		
2	Sales Order		U		R	U	U	U
	Production Order			U	C	U	U	U
	Individual Product					C	R	U
3	Shipping						C	
	Customer's Invoice		U					C

What is Information Lifecycle Management?

*Firstly, let’s understand what we mean by **Information Lifecycle***

The **Information Lifecycle** is the series of phases that information passes through, from inception, through use, to destruction.	**Information Lifecycle Management** is the means by which best practice is followed at each stage of the **Information Lifecycle**	**Information Lifecycle Management** defines the planning, implementation and control activities covering the operational, infrastructure, definition and functional areas of data management.

Information Lifecycle & SDLC



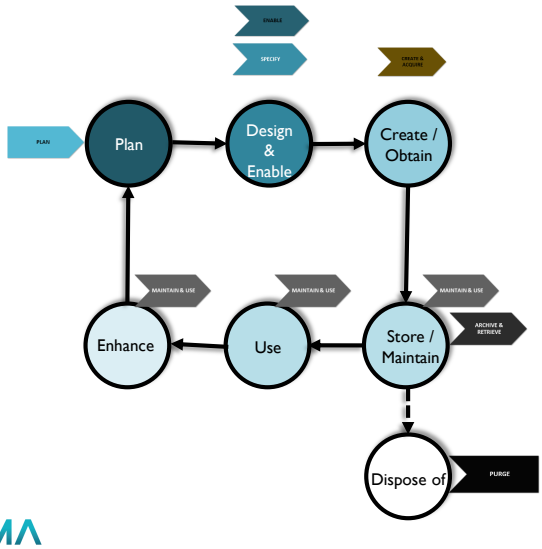
DMA
DATA MANAGEMENT ADVISORS

Data is precious

"Data is a precious thing and will last longer than the systems themselves."

Sir Tim Berners-Lee:
Inventor of the World Wide Web.

DMA
DATA MANAGEMENT ADVISORS



Data Lifecycle 4 "key" activities

1. Data Quality must be managed throughout the data lifecycle
2. Metadata Quality must be managed through the data lifecycle
3. Data Security must be managed throughout the data lifecycle
4. Data Management efforts should focus on the most critical data

What is the Information Lifecycle?

INFORMATION LIFECYCLE (DAMA)

PLAN	SPECIFY	ENABLE	CREATE & ACQUIRE	MAINTAIN & USE	ARCHIVE & RETRIEVE	PURGE
› *IM strategy* › *Governance* › *Define policies and procedures for quality, retention, security etc*	› *Architecture* › *Conceptual, logical and physical modelling*	› *Install or provision servers, networks, storage, DBMSs* › *Access controls*	› *Data created, acquired (external), extracted, imported, migrated, organised*	› *Data validated, edited, cleansed, converted, reviewed, reported, analysed*	› *Data archived, retained and retrieved*	› *Data deleted*

(SOURCE DAMA)

Information Lifecycle Activities

PLAN

- *Define information strategy*
- *Define documentation requirements*
- *Define governance framework*
- *Define monitoring framework*
- *Tool selection*
- *Define information classification framework*
- *Data retention and disposition plan*
- *Staff competencies and training map*
- *Create models for cost of data ownership*

The first stage of the lifecycle

(SOURCE DAMA)

DMA DATA MANAGEMENT ADVISORS

Information Lifecycle Activities

SPECIFY

- *Data requirements gathering*
- *Design data models*
  - *Conceptual*
  - *Logical*
  - *Physical*
  - *As is / To be*
- *Define taxonomies and ontologies*
- *Storage requirements*
- *Define backup requirements*
- *Version control systems*
- *Define warehousing requirements*

Data Modelling occurs in this lifecycle stage

(SOURCE DAMA)

DMA DATA MANAGEMENT ADVISORS

Information Lifecycle Activities

ENABLE

PROVISION/BUILD

- *Networks*
- *Servers*
- *Database Management Systems*
- *Data warehouses*
- *Embed performance monitoring*

APPLY ACCESS CONTROLS

- *Based on policies*
- *Build for availability*
- *Apply security controls*

Infrastructure is provisioned in this lifecycle stage

(SOURCE DAMA)

DMA DATA MANAGEMENT ADVISORS

Information Lifecycle Activities

CREATE & ACQUIRE

The lifecycle stage where information is acquired

- *Treat data as an asset*
- *Buy data from external sources*
  - *Postal addresses*
  - *Dunn and Bradstreet*
- *Creating reliable information*
  - *Data quality – how good is it?*
  - *Data currency – is it up-to-date?*
  - *Status and version control – is it draft or final?*
  - *Proof of provenance – who created it?*
- *Maintain creation information*

(SOURCE DAMA)

DMA DATA MANAGEMENT ADVISORS

Information Lifecycle Activities

MAINTAIN & USE

The lifecycle stage where business value is derived

- *Business value added during use*
  - *Supports business decisions*
- *Edited and updated*
  - *Data categorised and filed*
- *Monitor data quality, and fix*
- *Data may be transformed*
- *Accessed via data warehousing systems*
- *Used for reporting*
- *Data backed-up and restored*
- *Data distributed and disseminated*

(SOURCE DAMA)

DMA
DATA MANAGEMENT ADVISORS

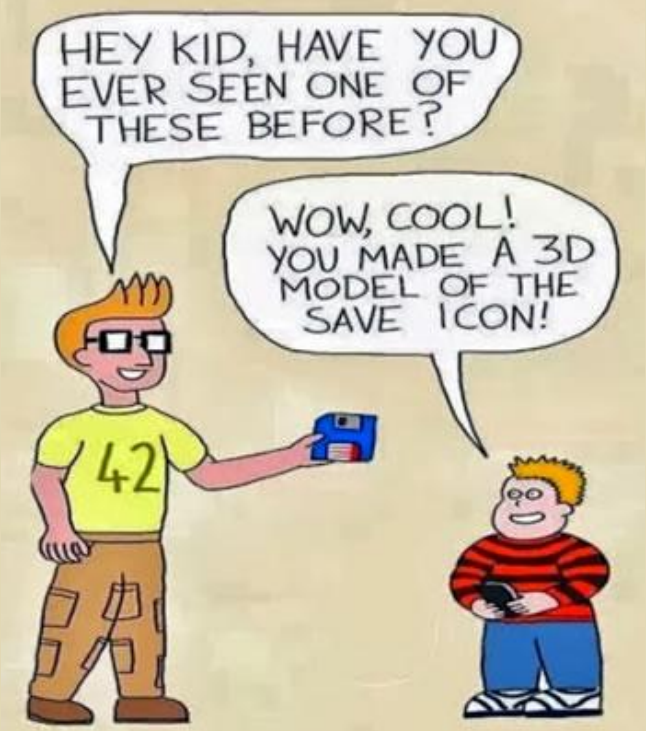
Information Lifecycle Activities

ARCHIVE & RETRIEVE

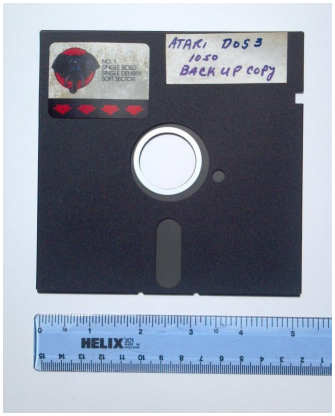
The lifecycle stage where information is archived (as specified during the planning stage)

- *Policy for archival periods*
  - *How long before data is archived?*
- *Test and verify retrieval methods*
- *Maintain and migrate archive formats*
  - *Media can degrade and become obsolete*
  - *Can old-format physical media still be read? (E.g. floppy disks)*
  - *Can legacy file formats still be read? (E.g. WordPerfect, Lotus 1-2-3)*
  - *Define procedures for validating archives*

DMA
DATA MANAGEMENT ADVISORS



DMA
DATA MANAGEMENT ADVISORS



DMA
DATA MANAGEMENT ADVISORS

Information Lifecycle Activities

PURGE

The lifecycle stage where information is deleted (as specified during the planning stage)

- *According to Retention Plan*
  - *Will vary by type of data*
- *Consider destruction method*
  - *Data is 'deleted' – may still be recoverable*
  - *Or physical destruction – shred CDs, hard drives, tapes, paper documents*
  - *"Secure Erase" for magnetic media*

(SOURCE DAMA)

DMA DATA MANAGEMENT ADVISORS

Key Points

- The Data Landscape is critical to monitor and govern data asset usage
- The Enterprise Data Model is the fundamental deliverable that:
  - Communicates Business Information Requirements
  - Provides the foundation & guidelines for Data Development
  - Aligns with other business models
  - Provides visual representation of business data
- Data Architecture includes architectural guidelines (reference) for the all the other data management capabilities
  - Operations, DW / BI, Meta-data, Integration, Meta-data, Quality
- The Data Lifecycle is flexible & tailorable for all types of technology

DMA DATA MANAGEMENT ADVISORS

Ref	Question	A	B	C	D	E
DA1	According to the DAMA DMBoK, what parts of the Data Life Cycle are integral parts of the SDLC?	Plan, Specify, Enable	Plan, Create & Acquire, Purge	Specify, Maintain & Use, Purge	Enable, Maintain & Use, Archive & Retrieve	Specify, Enable, Create & Acquire
DA4	Enterprise data architecture defines standard terms for things that are necessary to run an organization. These things are called:	Artefacts	Entities	Taxonomies	Meta-data	Relationships
DA5	The DMBOK identifies which of the following as common stages in the life cycle of the information asset	Plan, Obtain, Store/Share, Maintain, Apply, Dispose	Get, Store, Fix, Use, Purge	Acquire, Integrate, Apply, Share, Dump	Plan, Specify, Enable, Create and Acquire, Maintain & Use, Archive & Retrieve, Purge	Build/Buy, Mix/Merge, Apply, Delete
DA7	The key architecture domains include:	Zachmann, TOGAF, COBIT, and Heath architectures.	business, strategy, application and technology architectures.	business, data, infrastructure and technology architectures.	business, data, application and technology architectures.	process, database, software and technology architectures.

AFTER QUIZ 9

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)
8. Document, Records & Content Mgt (3)
9. Architecture & lifecycle (4)

Maximum possible score = 56

60% (CDMP Associate) = 34

70% (CDMP Practitioner) = 40

80% (CDMP Master) = 45

MetaData Management

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA DATA MANAGEMENT ADVISORS

Metadata Management (DMBoK 2 revised)

Definition: Planning, Implementation, and control activities that contributes to the ability to process, maintain, integrate, secure, audit and govern other data

Goals:
1. Provide organizational understanding of business terms and usage including technical lineage.
2. Collect and integrate metadata from diverse sources.
3. Provide a standard way to access metadata and enable known level of trust in data exchange.
4. Ensure metadata quality, consistency, currency, and security.

Business Drivers

Inputs:
- Business and Technical Metadata
- Operational Requirements
- Metadata Policy and Standards
- Data Architecture
- Metadata Issues
- Process Metadata
- Operational Metadata
- Data Governance Metadata

Activities:
1. **Define Metadata Strategy (P)**
2. **Understand Metadata Requirements (P)**
   1. Business User Requirements
   2. Technical User Requirements
3. **Define Metadata Architecture (P)**
   1. Create Metamodel (D)
   2. Apply Metadata Standards (C)
   3. Manage Metadata Stores (C)
4. **Create and Maintain Metadata (O)**
   1. Integrate Metadata (O)
   2. Distribute and Deliver Metadata (O)
5. **Query, Report and Analyze Metadata (O)**

Deliverables:
- Metadata Strategy
- Metadata Standards
- Metadata Architecture
- Metamodel
- Unified Metadata
- Metadata Stores
- Data Lineage
- Impact Analysis
- Metadata Control Process

Suppliers:
- Business Data Stewards
- Data Managers
- Data Governance Bodies
- Data Modelers
- Database Administrators

Participants:
- Data Stewards
- Project Managers
- Data Architects
- Business Analysts
- System Analysts

Consumers:
- Application Developers
- Data Analyst
- Data Integrators
- Business Users
- Knowledge Workers
- Customers & Collaborators
- Data Scientists
- Data Journalists
- Data Curators

Technical Drivers

Techniques:
- Data Lineage and Impact Analysis
- Metadata for Big Data Ingest

Tools:
- Metadata Repository Management Tools
- Metadata Repositories in other Tools
- Integration Tools

Metrics:
- Metadata Coverage Scorecard
- Metadata Quality Scorecard
- Metadata Repository Activities

(P) Planning, (C) Control, (D) Development, (O) Operations

DMA DATA MANAGEMENT ADVISORS

What is Metadata?

DMA
DATA MANAGEMENT ADVISORS

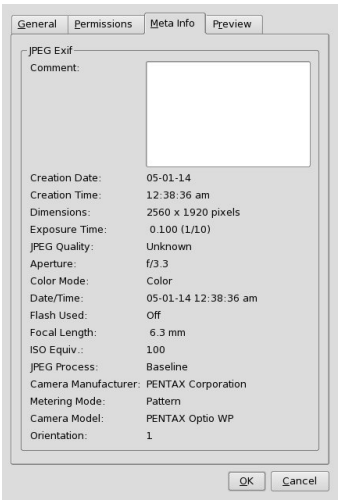
Where do YOU encounter metadata every day?

DMA
DATA MANAGEMENT ADVISORS

MetaData

DATA

METADATA

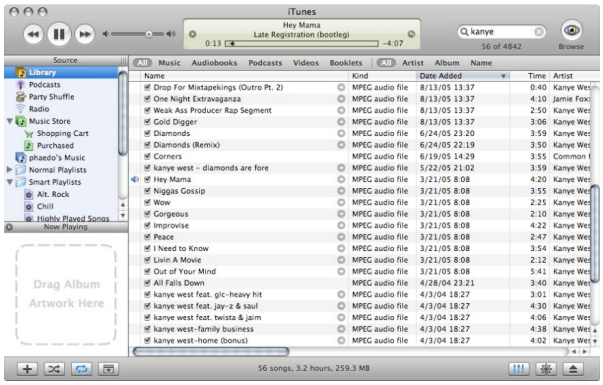


DMA
DATA MANAGEMENT ADVISORS

MetaData

DATA

METADATA



DMA
DATA MANAGEMENT ADVISORS

Why Do We Need Metadata?

DMA DATA MANAGEMENT ADVISORS

Metadata adds Context and Definition

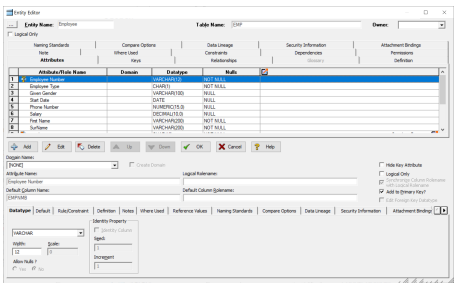
Is this the city where the customer lives
or
the city where the store is located?

First Name	Last Name	Company	City	Year Purchased
Kimi	Raikkonen	Komputers R Us	Helsinki	1970
Jenson	Button	The Lord's Store	London	1999
Ayrton	Senna	The Lady's Store	Sao Paulo	1998
Sebastian	Vettel	My Favourite Store	Berlin	2001

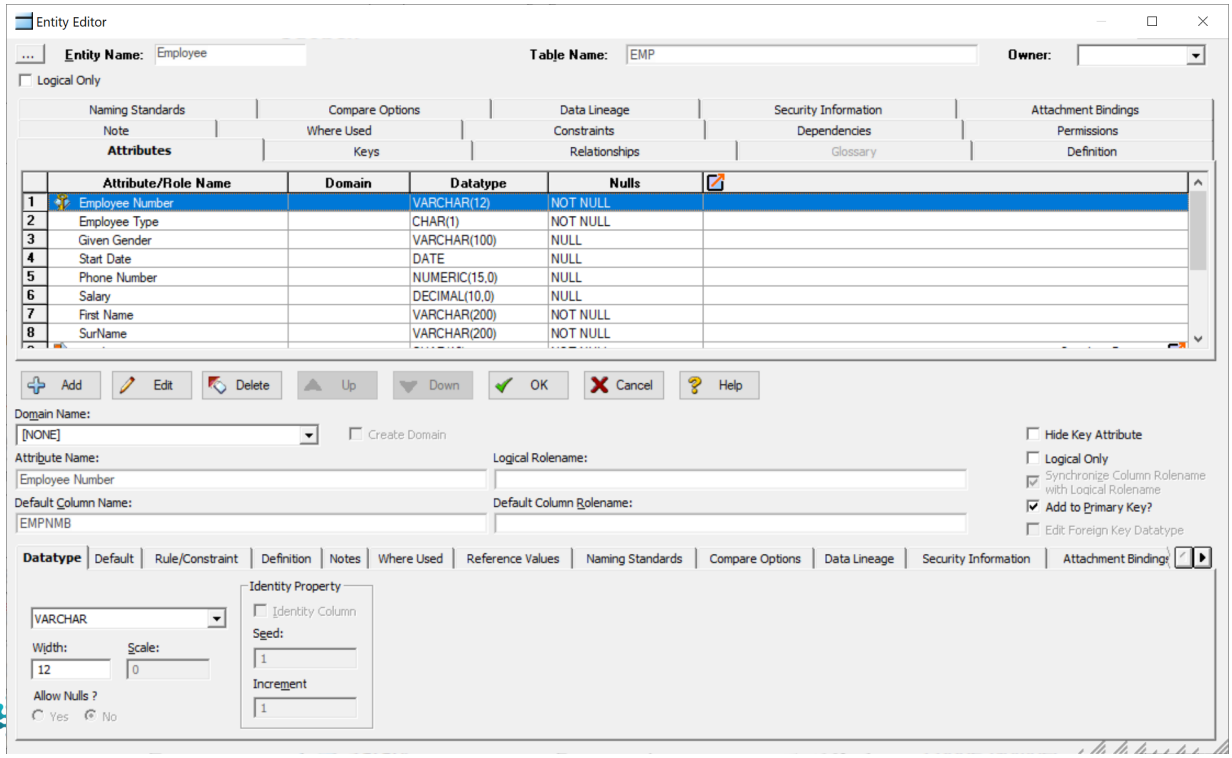
Properties (metadata) of the attributes

Definition	Last Name represents the surname or family name of an individual.
Business Rules	In the Chinese market, family name is listed first in salutations.
Format	VARCHAR(30)
Abbreviation	LNAME
Required	YES
Etc.	Numerous technical & business metadata including security, privacy, nullability, primary key, etc.

Beware – verylittle (to zero) metadata with big data technologies



DMA DATA MANAGEMENT ADVISORS



Metadata covers the 6 interrogatives of Data

Who	What	Where	Why	When	How
Who is regulating or auditing this data?	What are the technical naming standards for database implementation?	Are there local privacy or security policies that regulate this data?	Why are we regulating or auditing this data?	When do regulators and / or auditors require the data?	How are the auditors / regulators permitted to request this data?
Who created this data?	What is the business definition of this data element?	Where is this data stored?	Why are we storing this data?	When was this data created?	How is this data formatted? (character, numeric, etc.)
Who is the Steward of this data?	What are the business rules for this data?	Where did this data come from?	Why are we storing this data ie its usage & purpose?	When was this data last updated?	How many databases or data sources store this data?
Who is using this data?	What is the security level or privacy level of this data?	Where is this data used & shared?	Why is this data managed, i.e. the business motivations?	When does it need to be purged / deleted?	How / by what mechanisms / devices is this data handled?
Who "owns" this data?	What is the abbreviation or acronym for this data element?	Where is the backup for this data?	Why is the data classified as PII?	When should the data be updated?	How is the data provided to the regulators?
Who is permitted to use this data?	What are the business drivers for using this data?	Where is the disaster recovery location for this data?	Why is this data business critical?	When was the data last quality checked?	How long should it be stored?

DMA DATA MANAGEMENT ADVISORS

Automatic capture | Manual capture | Mixed capture

Types & typical Sources of MetaData

- Application Code
- Big Data platforms (but not much here)
- Business Intelligence (BI) Tools
- Business Process Models
- CMDB
- COBOL copybooks
- Data Models (extended metadata)
- Data Transformation (e.g. ETL Tools)
- Data Quality Tools
- ERP, CRM, and Package Applications
- Humans
- Internet of Things (IoT)
- LDAP / IDAM
- Open Data
- Photos / Images (EXIF)
- PREMIS (OAIS Reference Model (ISO 14721) Open Archival Information System (or OAIS) PREservation Metadata: Implementation Strategies (PREMIS)
- Relational databases (system catalogue)
- Social Media
- Spreadsheets
- Text Documents
- Transaction logs
- XML
- ….. and more

DMA DATA MANAGEMENT ADVISORS

NoSQL – Key Value Databases

4 types of NoSQL databases:
1. Document
2. Key Value
3. Graph
4. Columnar

- NoSQL Databases are often optimal solutions for flexibility & performance in certain scenarios.
  - One common NoSQL database is a key-value pair database (e.g. Redis, RIAK, Oracle NoSQL, etc.)
  - They can support extremely high volumes of records & state changes per second through distributed processing and distributed storage.
  - Use cases include: Managing user sessions in web applications, online gaming, online shopping carts, etc.

Key	Value
1839047	John Doe, Prepaid, 40.00
9287320	01/01/2008, 50.00, Green

  - The structure is often created by the **application code**, not within a database or metadata structure.
    - Metadata for NoSQL databases is typically minimal or non-existent.
    - The structure & metadata is generally determined by the application code

DMA DATA MANAGEMENT ADVISORS

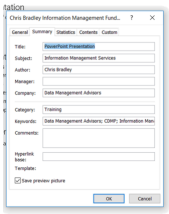
Types & typical Sources of MetaData

The OCCURS clause marks an array.

The DEPENDING ON clause marks a counter field for the array, if one exists.

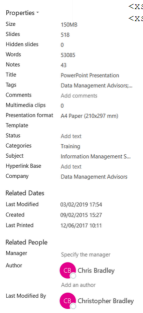
The array ASSIGMENTS is a nested array within COURSES.

```
01  STUDENT.
    20  ID                          PIC 9(8).
    20  FIRST_NAME                  PIC X(32).
    20  LAST_NAME                   PIC X(32).
*                                   YYYYMMDD
    20  DATE_OF_BIRTH               PIC S9(8) COMP.
    20  NUMOF_COURSES               PIC 9(4) COMP.
    20  NUMOF_BOOKS                 PIC 9(4) COMP.
    20  COURSES.
      25 COURSE OCCURS 8 TIMES DEPENDING ON NUMOF_COURSES.
        30  COURSE_ID               PIC 9(8).
        30  COURSE_TITLE            PIC X(48).
        30  INSTRUCTOR_ID           PIC 9(8).
        30  NUMOF_ASSIGNMENTS       PIC 9(4) COMP.
        30  ASSIGNMENTS OCCURS 4 TIMES DEPENDING ON NUMOF_ASSIGNMENTS.
          40  ASSIGNMENT_TYPE       PIC X(12).
          40  ASSIGNMENT_TITLE      PIC X(48).
*                                   YYYYMMDD
          40  DUE_DATE              PIC S9(8) COMP.
          40  GRADE                 PIC S9V9.
    20  BOOKS.
      25  BOOK OCCURS 1 TO 5 TIMES DEPENDING ON NUMOF_BOOKS.
        30  ISBN                    PIC X(10).
*                                   YYYYMMDD
        30  RETURN_DATE             PIC 9(8) COMP.
```



BookCatalog.xsd

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <xsd:element name="BookCatalog">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Book" minOccurs="1"
  maxOccurs="unbounded"/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>
    <xsd:element name="Book">
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Title" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Author" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Date" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="ISBN" minOccurs="1" maxOccurs="1"/>
                <xsd:element ref="Publisher" minOccurs="1" maxOccurs="1"/>
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>
    <xsd:element name="Title" type="xsd:string"/>
    <xsd:element name="Author" type="xsd:string"/>
    <xsd:element name="Date" type="xsd:string"/>
    <xsd:element name="ISBN" type="xsd:string"/>
    <xsd:element name="Publisher" type="xsd:string"/>
```



Social Media Metadata

Metadata from social media, such as Twitter, can help identify trend and sentiment analysis.

- Author
- Text/Content of Tweet
- Hash Tag
- # of Retweets

- Date/Time.
- Location.
- Language.
- Device / OS.
- User ID.
- ID of Followers.
- ID of users who "liked" tweet.
- ID of users retweeting tweet.
- Spaces.
- Direct Messages.
- Lists.
- Trends.
- Media.
- Places.
- Etc.

Types of metadata 3 CATEGORIES - DMBoK

Categories in relation to where it originates vs where it is used

Business metadata

Relates business perspective to the metadata user (e.g., business data definitions, regulatory or contractual constraints, Data Quality statements).

Examples include:

- Definitions and descriptions of data sets, tables, and columns
- Business rules, transformation rules, calculations, and derivations
- Data models
- Data quality rules and measurement results
- Schedules by which data is updated
- Data provenance and data lineage
- Data standards
- Designations of the system of record for data elements
- Valid value constraints
- Stakeholder contact information (e.g., data owners, data stewards)
- Security/privacy level of data
- Known issues with data
- Data usage notes

Technical

The technical details of the data in IT systems (e.g., systems that store data, processes that move and use it)

Examples include:

- Physical database table and column names
- Column properties
- Database object properties
- Access permissions
- Data CRUD (create, read, update and delete) rules
- Physical data models, including data table names, keys, and indexes
- Documented relationships between the data models and the physical assets
- Data Integration / loading job details
- File format schema definitions (DDL)
- Source-to-target mapping documentation
- Data lineage documentation, including upstream and downstream change impact information
- Program and application names and descriptions
- Content update cycle job schedules and dependencies
- Recovery and backup rules
- Data access rights, groups, roles

Operational metadata

Targeted at IT operations users' needs (e.g., data archiving and retention rules, audit rules, purge criteria)

Examples include:

- Logs of job execution for batch programs
- History of extracts and results
- Schedule anomalies
- Results of audit, balance, control measurements
- Error Logs
- Reports and query access patterns, frequency, and execution time
- Patches and Version maintenance plan and execution, current patching level
- Backup, retention, date created, disaster recovery provisions
- SLA requirements and provisions
- Volume and usage patterns
- Data archiving and retention rules, related archives
- Purge criteria
- Data sharing rules and agreements
- Technical roles and responsibilities, contacts

DMA DATA MANAGEMENT ADVISORS

Types of metadata – Others (Real World)

- **Process metadata**
  Other system elements e.g.
  - Process name,
  - Data Stores & Data Involved,
  - Government/ Regulatory Bodies,
  - Roles & Responsibilities,
  - Process Handoffs,
  - Process Dependencies
  - Decomposition …
- **Data Stewardship / Governance metadata**
  Data about stewards and stewardship processes (e.g.
  - Data owners,
  - Data subject areas,
  - Data users, Data stewards,
  - Data Classification …

DMA DATA MANAGEMENT ADVISORS

Types of metadata – Information Science

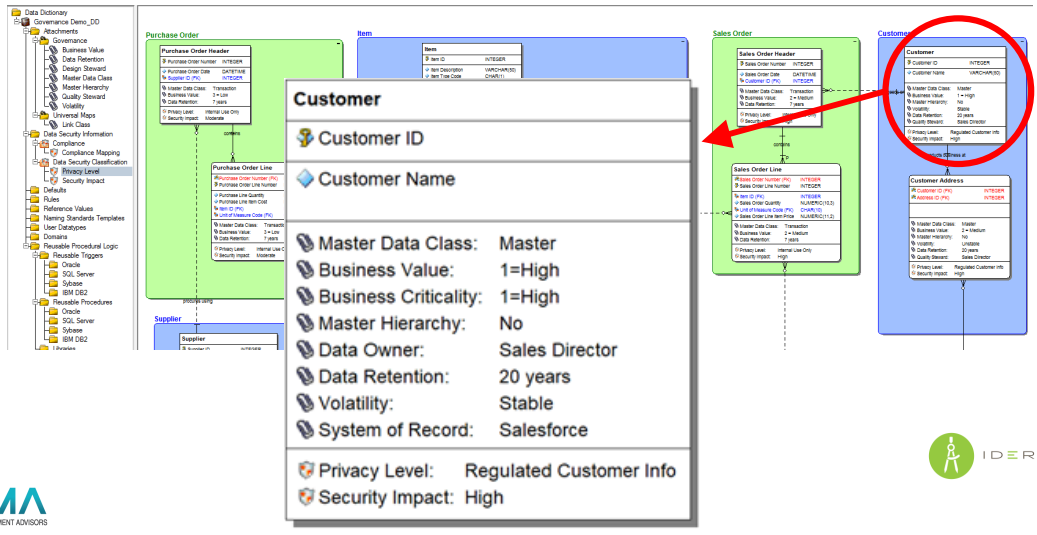
Separate to IT i.e. information science, Metadata is described using 3 different categories:

- **Descriptive Metadata**
  e.g., title, author, reference #, subject, keywords. This describes a resource and enables identification and retrieval.
- **Structural Metadata**
  This describes relationships within and among resources and their component parts e.g., number of pages, number of chapters, index, references, etc.
- **Administrative Metadata**
  e.g., version numbers, archive dates, last updated, approvers etc. This is used to manage resources over their lifecycle – essential for some Document Control Schemes (DCS)

DMA DATA MANAGEMENT ADVISORS

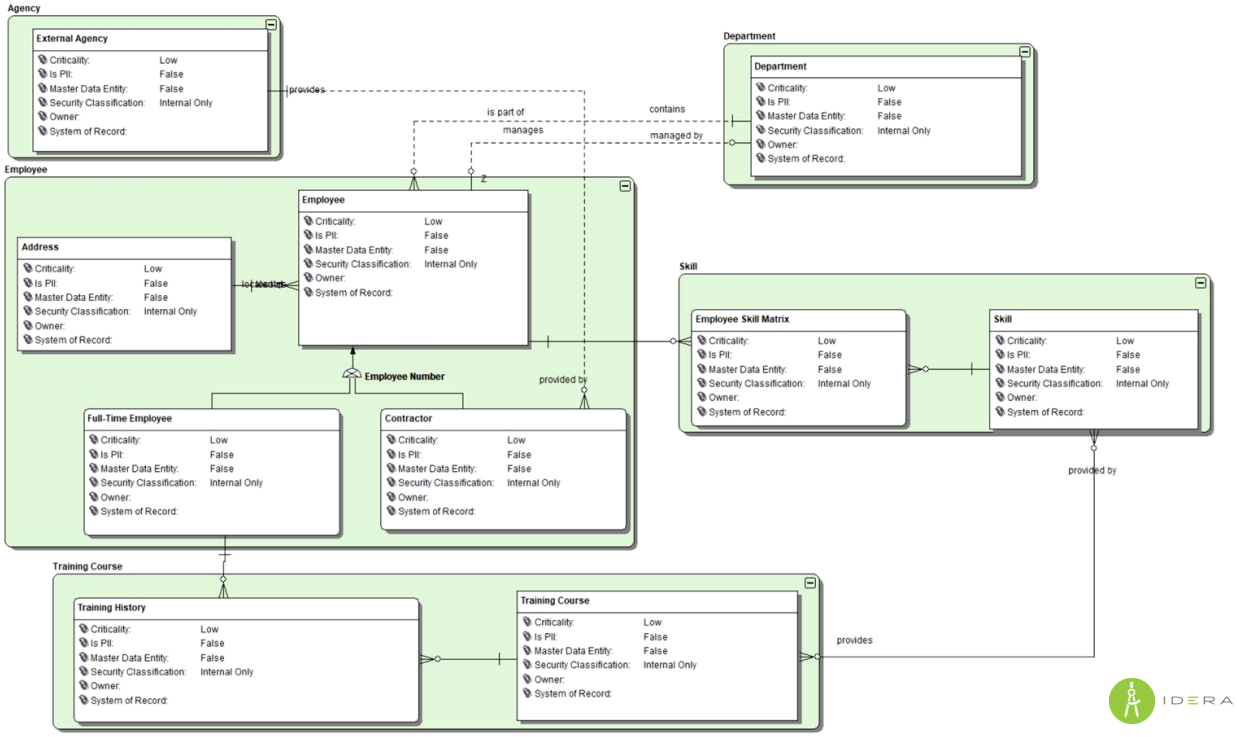
Data Governance & Models – Metadata extensions

(ER/Studio Attachments)



IDERA

DMA DATA MANAGEMENT ADVISORS



Metadata & Data Governance

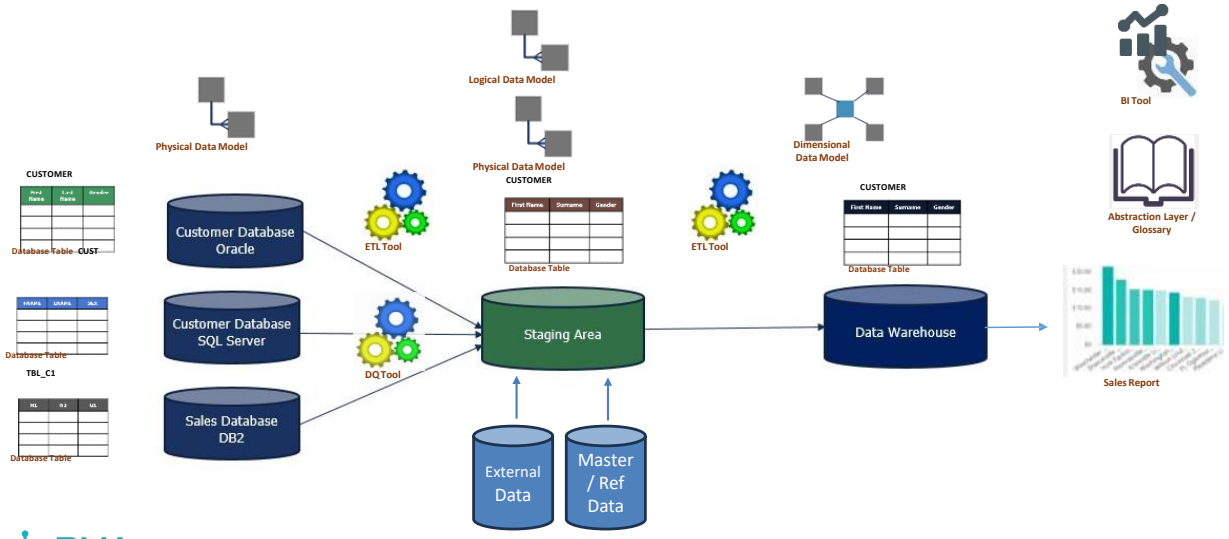
Major data governance aspects held in metadata include:

- **Business Glossary:** Defining the business metadata definitions for business data elements.
- **Data Dictionary:** Defining the technical metadata definitions for data(base)objects (fields, files, tables etc)
- **Data Stewardship:** Aligning data stewardship or ownership roles to key data objects.
- **Data Standards:** Metadata standards provide governance and rules for current and future development, based on both business and IT input.
- **Privacy & Security:** Identification of privacy and security levels for business data is a key aspect of data governance can be managed with metadata definitions.
- **Traceability & Audit:** Understanding of how data is used across the organization, how key financial figures are calculated, etc.

Metadata is a key enabler for a Data Governance initiative

Data Lineage

Data Warehousing Example …. Metadata for CUSTOMER exists in a number tools & data stores



Data Values Standards (some examples)

COUNTRY Codes:

- **ISO 3166**
  - Alpha-2, Apha-3, Numeric codes for countries, codes for subdivisions and formerly used codes
- **FIPS 10**
- **IOC**
- **Licence Plate**
- **Internet Domain (ICANN)**

CURRENCY Codes:

- **ISO 4217**
  - **ISO 3166** Alpha-2 + 1 char Currency
  - E.g. GBP, USD, CHF

DATE / TIME:

- **ISO 8601**
  - D = yyyymmdd
  - T= hhmmss

LANGUAGE Codes:

- **ISO 639**
  - Alpha-2, Alpha-3, Name, Scope, Type
  - EN, ENG, English, Individual, Living
  - LA, LAT, Latin, Individual, Ancient

UNIT OF MEASUREMENT (UOM):

- **7 SI base units**
  - Length - meter (m)
  - Time - second (s)
  - Amount of substance - mole (mole)
  - Electric current - ampere (A)
  - Temperature - kelvin (K)
  - Luminous intensity - candela (cd)
  - Mass - kilogram (kg)

CLASSIFICATION OF DISEASES

- World Heath Organisation (WHO) International Classification of Diseases (ICD). The standard for all clinical & research purposes.

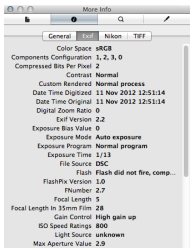
Country	FIPS10-4	ISO 3166
Aland Islands	–	AX
Algeria	AG	DZ
American Samoa	AQ	AS
Antigua & Barbuda	AC	AG
Australia	AS	AU
Austria	AU	AT
Bahrain	BA	BH
Bosnia and Herzegovina	BK	BA
Gabon	GB	GA
Saudi Arabia	SA	SA
South Sudan	-	SS
Sudan	SU	SD
United Kingdom of Great Britain and Northern Ireland	UK	GB

Data Values Standards (some examples)

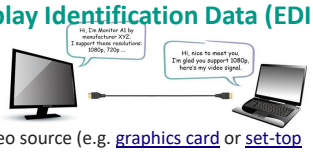
DATES: ISO 8601

Dates constructs in ISO 8601	
Date	2021-10-23
Date and time in UTC	2021-10-23T15:08:07+00:00 2021-10-23T15:08:07Z 20211023T150807Z
Week	2021-W42
Week with weekday	2021-W42-6
Date without year	--10-23
Ordinal date	2021-296

Exchangeable image file format (officially **Exif**, according to JEIDA/JEITA/CIPA specifications) 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)[2] 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 μ-Law PCM for uncompressed audio data, and IMA-ADPCM for compressed audio data).[3] It is not used in JPEG 2000 or GIF. This standard consists of the Exif image file specification + Exif audio file specification.

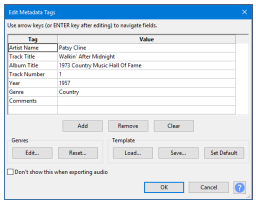


Extended Display Identification Data (EDID)



A metadata format for display devices to describe their capabilities to a video source (e.g. graphics card or set-top box). The data format is defined by a standard published by the Video Electronics Standards Association (VESA). The EDID data structure includes manufacturer name and serial number, product type, phosphor or filter type, timings supported by the display, display size, luminance data and (for digital displays only) pixel mapping data.
DisplayID is a VESA standard targeted to replace EDID and E-EDID extensions with a uniform format suited for both PC monitor and consumer electronics devices

ID3 is a metadata container most often used in conjunction with the MP3 audio file format. It allows information such as the title, artist, album, track number, and other information about the file to be stored in the file itself. There are two unrelated versions of ID3: ID3v1 and ID3v2. ID3v1 takes the form of a 128-byte segment at the end of an MP3 file containing a fixed set of data fields. ID3v1.1 is a slight modification which adds a "track number" field at the expense of a slight shortening of the "comment" field. ID3v2 is structurally very different from ID3v1, consisting of an extensible set of "frames" located at the start of the file, each with a frame identifier (a three- or four-byte string) and one piece of data. 83 types of frames are declared in the ID3v2.4 specification, and applications can also define their own types. There are standard frames for containing cover art, BPM, copyright and license, lyrics, and arbitrary text and URL data.



Industry Metadata Standards

- OMG (Common Warehouse Metadata (CWM),
- Information Management Metamodel (IMM),
- MDC Open Information Model (OIM), XML, UML, SQL)
- World Wide Web Consortium (W3C): RDF (Relational Definition Framework)
- Dublin Core: Dublin Core Metadata Initiative (DCMI)
- Distributed Management Task Force (DTMF): Web-Based Enterprise Management (WBEM)
- XML Metadata Interchange (XMI)
- Metadata Standards for Unstructured Data
- ISO / IEC 11179 Metadata Registry Standard for naming of concept, data element concept, conceptual domain, data element, and value domain.
- Data Catalog Vocabulary (DCAT) from W3C

CWM defines a specification for modelling metadata for relational, non-relational, multi-dimensional, and most other objects found in a data warehousing environment. Uses UML, MOF & XMI

Dublin Core Schema is a small set (15) of vocabulary terms that can be used to describe digital resources (video, images, web pages, etc.), as well as physical resources such as books or CDs and objects like artworks.

OMG standard for exchanging metadata information via XML. Current version 2.5.1

ISO metadata registries (MDR) standard

Meta-Object Facility (MOF) An Object Management Group standard for model-driven engineering. Its purpose is to provide a type system for entities in the CORBA architecture and a set of interfaces through which those types can be created and manipulated.

Metadata & EDI

For EDI (Electronic Data Interchange), the primary metadata standard is UN/EDIFACT, an international standard developed by the United Nations, widespread outside of North America;

Within North America, the most common EDI metadata standard is ANSI ASC X12 (X12).

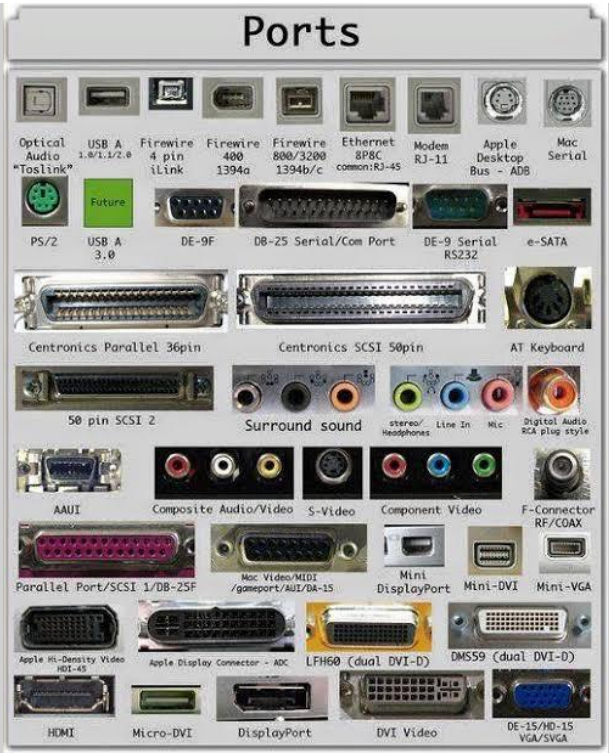
- **Data element structure:**

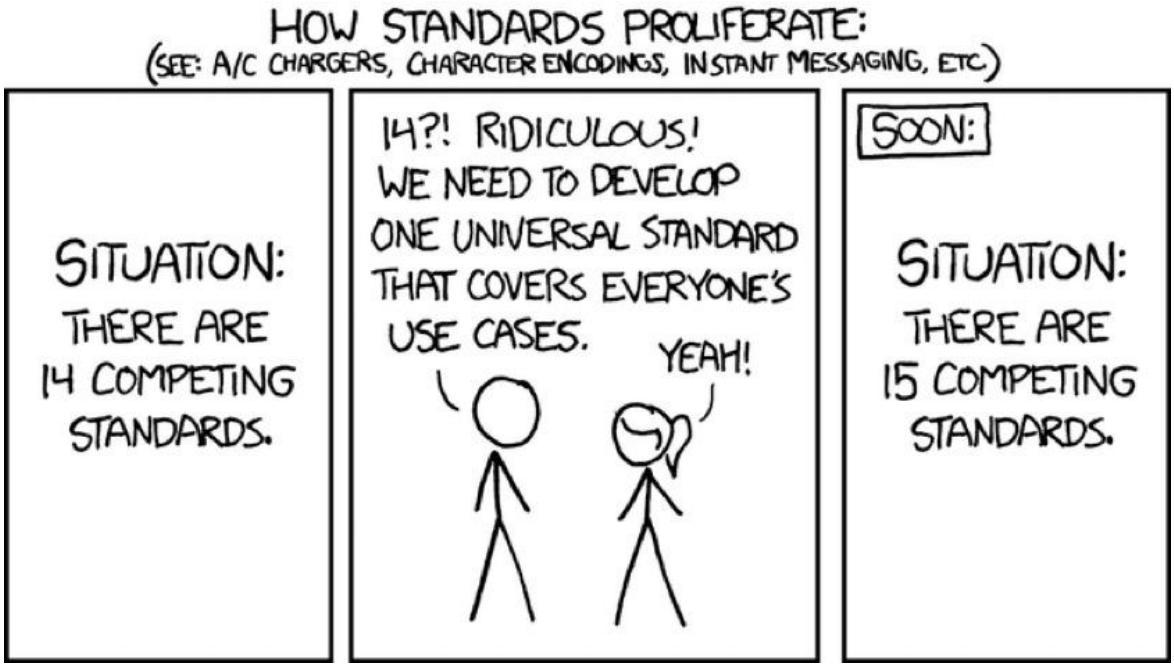
  EDI standards define the structure of data elements within a message, including data types, codes, and lengths, ensuring consistency in data exchange.
- **Message types:**

  Each EDI standard includes various message types, such as purchase orders, invoices, shipping notices, which specify the required data fields for that particular transaction.
- **Industry-specific subsets:**

  In addition to the primary standards, different industries have their own subsets of EDI standards tailored to their needs, e.g.
  - EANCOM: retail
  - TRADACOMS: Primarily used in the UK retail sector
  - ODETTE: automotive industry.
  - VDA: European automotive industry, mainly in Germany
  - HL7: A semantic interoperability standard commonly used in healthcare data exchange

DMA
DATA MANAGEMENT ADVISORS





Ref	Question	A	B	C	D	E
MM1	The role of the Physical data model in the Metadata repository is	Which version of COTS software (E.g. SAP) is implemented	To describe how and where our data is stored in our systems applications or packages.	What the business definition of data concepts is	How many master data records are stored in our MDM system	When the duplicated records were merged
MM2	Updating the Metadata repository is a recommended activity during project close out in the SDLC	TRUE	FALSE			
MM3	What type of Meta-Data provides developers and administrators with knowledge and information about systems?	Unstructured Meta-Data	Process Meta-Data	Business Meta-Data	Data Stewardship Meta-Data	Technical Operational Meta-Data
MM4	These are examples of which type of Meta-Data: Data Stores & Data Involved, Government/ Regulatory Bodies; Roles & Responsibilities; Process Dependencies and Decomposition	Technical Meta-Data	Business Meta-Data	Process Meta-Data	Data Stewardship Meta-Data	Operational Meta-Data
MM5	Which of the following is a Meta-Data scheme focused specifically on documents?	Administrative Meta-Data	Structural Meta-Data	Descriptive Meta-Data	Preservation Meta-Data	Business Meta-Data
MM6	Which of the following contain metrics associated with Meta-Data Management	Steward Representation/ Coverage; Meta-Data Repository Availability; Meta-Data Management Maturity	Meta-Data Management Maturity; No. of Data Stewards; No. of Meta-Data Attributes Listed	Steward Representation/ Coverage; Meta-Data Repository Availability; No. external Reference Data Sources	Meta-Data Repository Availability; No. Data Governance Meetings held annually; No. Meta-Data Repositories	No. of Entities, Attributes & Relationships stored in the Repository

AFTER QUIZ 10

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)
8. Document, Records & Content Mgt (3)
9. Architecture & lifecycle (4)
10. Metadata management (6)

Maximum possible score = 62

60% (CDMP Associate) = 38

70% (CDMP Practitioner) = 44

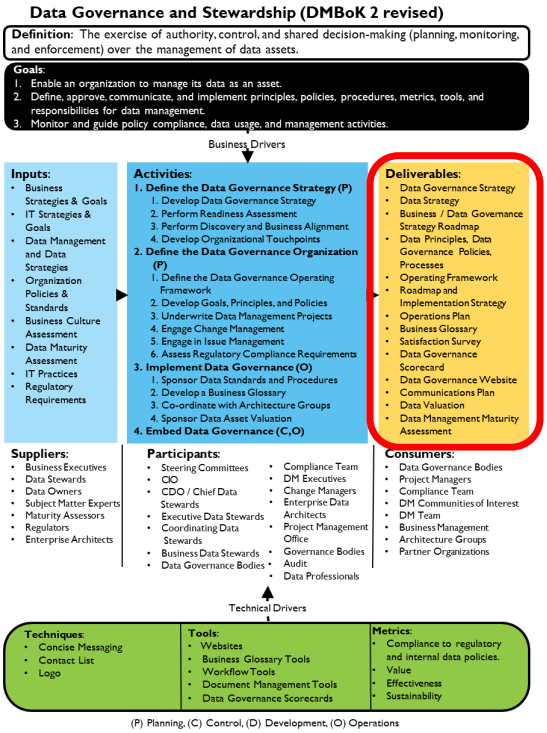
80% (CDMP Master) = 50

Data Governance

.... at the *Heart* of ALL
Information Management Disciplines

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA
DATA MANAGEMENT ADVISORS



DMA DATA MANAGEMENT ADVISORS

Other Definitions of Data Governance

"Data Governance is a quality control discipline for adding new rigor and discipline to the process of managing, using, improving and protecting organizational information."
(IBM Data Governance Council)

"The process of managing and improving data for the benefit of all stakeholders"
(Chris Bradley)

"Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."
(Data Governance Institute)

"Data Governance is the formal orchestration of people, processes, and technology to enable an organization to leverage data as an enterprise asset."
(MDM Institute)

DMA DATA MANAGEMENT ADVISORS

What Is Data Governance?

The Design & Execution Of Standards & Policies Covering …

- Design and operation of a management system to assure that data **delivers value**
- Who can do what to the organisation's data and how
- Ensuring standards are set and met
- A strategic & high-level view **across the whole organisation**

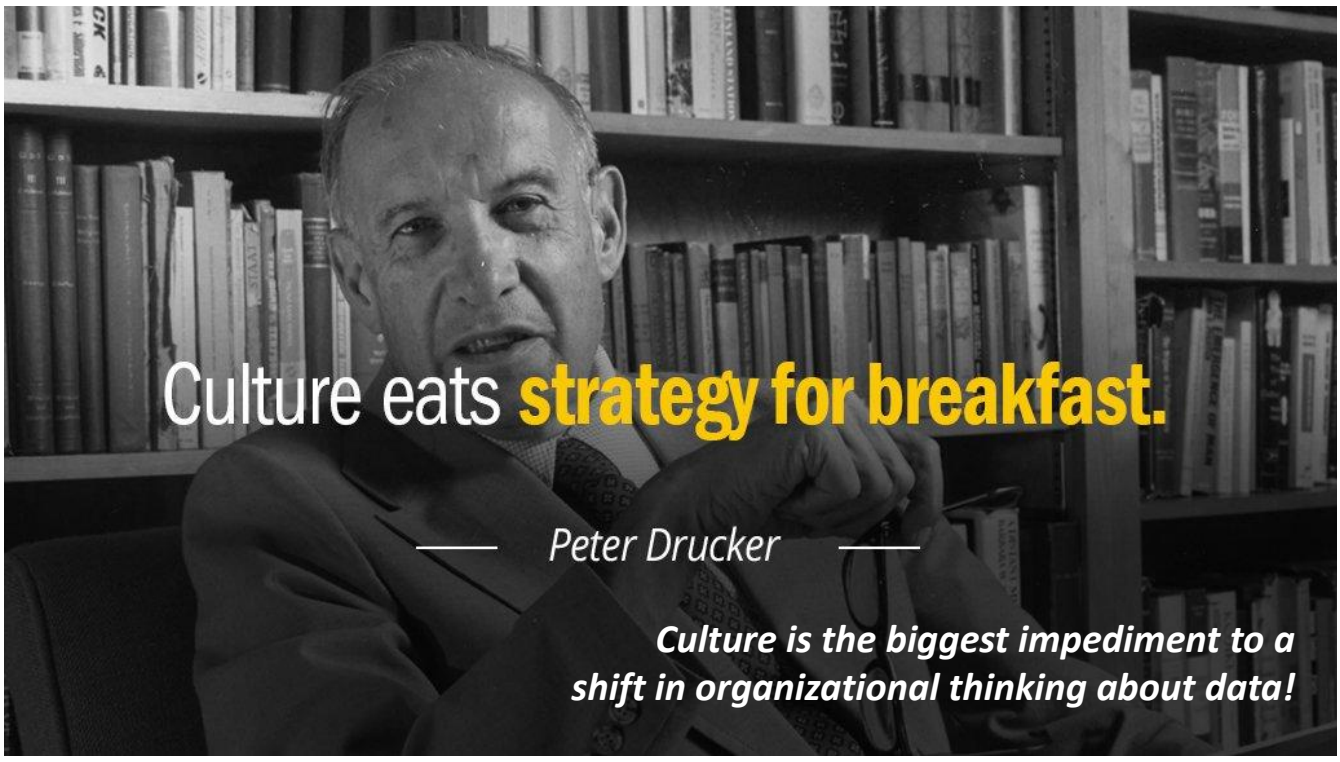
To Ensure …

- Key principles/processes of effective Information Management are put into practice
- Continual improvement via evolution of an Information Management strategy
- Management of Data as a Corporate asset becomes part of the **culture**

Data Governance Is NOT …

- A "one off" Tactical management exercise
- The responsibility of the Technology and IT department alone

DMA
DATA MANAGEMENT ADVISORS



Assets

ISO 55000 defines Asset Management as:

the coordinated activity of an organisation to realize value from assets.

ISO 55000 defines Assets as:

An asset is an item, thing or entity that has potential or actual value to an organisation.

DMA
DATA MANAGEMENT ADVISORS

Data as an asset

Data is an asset that has value and should be managed and measured accordingly throughout its lifecycle

DMA
DATA MANAGEMENT ADVISORS

DATA is our ONLY

- NON Depletable
- Reusable
- Copyable
- Enterprise wide

Asset

Key Points

- Data Governance is core (foundational) function of data management
- Data Stewardship comprises business data stewards with data management professionals (IT)
- A Data Governance program requires organizational structure and activities
- Data Governance is a ongoing program

Data Governance:
A Core DM Function

- Data Governance interacts and guides how all other data management functions are performed.
- Requires understanding of business strategies and objectives including regulatory constraints to set effective data strategy for all functions.
- Data Governance has a role in the data management activities of all business projects in an enterprise.

Data Stewardship

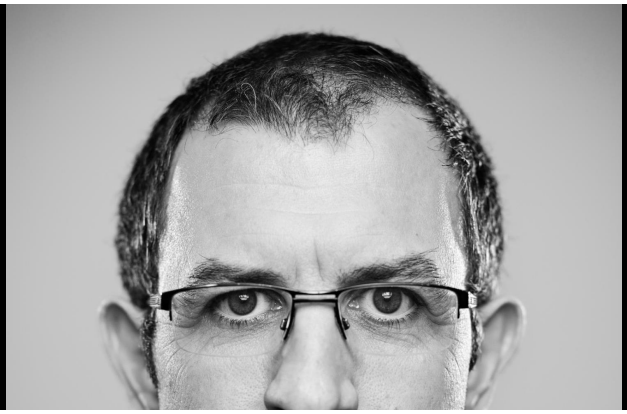
- Comprised of business data stewards and subject matter experts (SMEs) partnered with Data Management professionals (IT) that are accountable for the effective management and use of organization's enterprise data assets.
- *Typical* data stewards' types include:
  - **Executive** – Senior managers serve on Data Governance Steering Committee & DGC (A "Chief Data Steward" or CDO if you have one, typically chairs the Data Governance Steering Committee (DGSC).
  - **Coordinating** – Leads team of business data stewards.
  - **Business** – Recognized data subject matter experts in organization.
- Data stewards' responsibilities can span across all Data Management functions.

DMA DATA MANAGEMENT ADVISORS

Type of Data Stewards (DMBoK 2_R)

- **Executive Data Stewards** are senior managers who serve on a Data Governance Council.
- **Enterprise Data Stewards** have oversight of a data domain across business functions.
- **Technical Data Stewards** are IT professionals operating within one of the Knowledge Areas, such as Data Integration Specialists, Database Administrators, Business Intelligence Specialists, Data Quality Analysts, or Metadata Administrators.
- **Chief Data Stewards** may chair Data Governance bodies in lieu of the CDO or may act as a CDO in a virtual (committee-based) or distributed Data Governance organization. They may also be Executive Sponsors.
- **Coordinating Data Stewards** lead and represent teams of business and technical Data Stewards in discussions across teams and with executive Data Stewards. Coordinating Data Stewards is particularly important in large organizations..
- **Business Data Steward**
  Business Data Stewards are business professionals, most often recognized subject matter experts, accountable for a subset of data. They work with stakeholders to define and control data.
- **A Data Owner** is a businessperson who is accountable for decisions about data within their domain.

DMA DATA MANAGEMENT ADVISORS



Data Owner

- **Accountable** for the Data Subject area for which they are the "owner"
- Work closely with Data Stewards to establish consistent data quality and business rules
- Define & approve quality, access and security requirements.
- Ensure data for the data subject area is fit for purpose
- Communication & Stakeholder management skills are essential
- Corporate **Accountability** for the Data subject area …. must **not** simply focus on "their" departmental needs.

DMA
DATA MANAGEMENT ADVISORS

Data Steward

- Is a ***business*** role appointed to take responsibility for the quality and use of their organization's data assets.
- Develop and maintain definitions
- Authority for data and metadata definitions
- Create ongoing business rules for data
- Support business analysts in the alignment of functional processes and data requirements
- Assist data quality manager in defining metrics, matching and standardisation rules
- Track and determine the need for additional data elements on projects and opportunities for re-use across projects and business processes
- Is a recognized subject matter expert in the data subject area / business domain that he or she is responsible for
- Works collaboratively across the organization with data stakeholders and others Identifying data problems & issues
- Is an effective communicator
- Works in association with the Data Owner to protect and enhance the data assets under his or her control

DMA DATA MANAGEMENT ADVISORS

Data Steward - Role Activities

1. Ensuring data is managed in accordance with IMF and data process exist which others can follow
2. Implementing appropriate controls and data governance mechanisms to ensure data is fit for purpose, based on agreed process/lifecycle
3. Defining and monitoring key data to agreed standards and quality (business and regulatory (e.g. BASEL)) needs
4. Assisting in understanding the business data needs to ensure data is fit for all customers (data mapped to processes)
5. Monitoring & enforcing data policies & practices in their business function
6. Communicating & promoting the value of IMF, to ensure data is right first time through appropriate IMF roles
7. Reviewing and monitoring data quality analysis & audits, including MI/outputs from data quality tools
8. Assist in data quality analysis & improvement and prioritise data for remediation
9. Data Triage - Identify and remediate data issues within own business unit or across the Enterprise
10. Providing first point of contact to resolve IM issues
11. Providing input to Data policies, standards & procedures for business function and across the Enterprise
12. Defining and agreeing of business data quality rules for data quality tool

DMA DATA MANAGEMENT ADVISORS

Assist the Data Owners in the implementation & adherence to:

i. Information Management Principles
ii. Global Data Privacy Policy
iii. Global Information and Records Management Procedure
iv. Group Policy Information Security Management
v. Group Policy Protection and Disclosure of Price Sensitive Information
vi. Group Policy Protection and Disclosure of Personal (PII) Information

Other responsibilities

a) Openly advocating the benefits of a comprehensive Information Management Framework
b) Educating others on and assisting to implement policies, processes, standards, and procedures
c) Identifying opportunities to better manage information across functions
d) Defining, implementing and reviewing appropriate Information Quality standards
e) Accepting responsibility for the implementation of information quality performance measures
f) Assisting with gathering and collating feedback for Steering Groups and Data Owners
g) Collaborating with the wider Information Management community to ensure information quality procedures are correctly designed and followed
h) Managing the demand for changes to processes, policies, and procedures against budget and resource constraints
i) Facilitating the prioritisation of work, scheduling and assignment of information quality initiatives
j) Resolving day-to-day process issues in coordination with other Data Stewards and departmental managers
k) Assisting with the development training plans, materials and coordinating of training activities related to Information Management

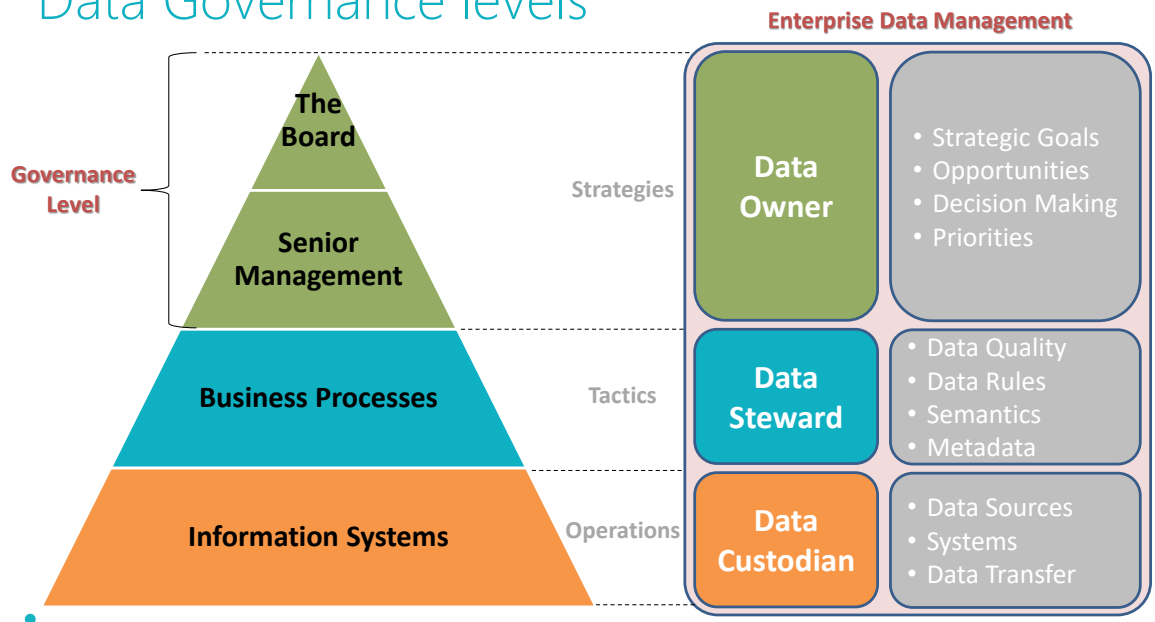
Data Custodian

- Provide formats for operational data and ensure reused
- Assist Data Stewards in interpretation and definition
- Profile and explain source system details
- Work with DQ Manager to correct data at the source
- Explain operational systems processing
- Attend DGC to provide expertise on data sources, targets and transformation processes
- Ensure data is backed up, archived & available to the Business

Container vs Content

DMA DATA MANAGEMENT ADVISORS

Data Governance levels



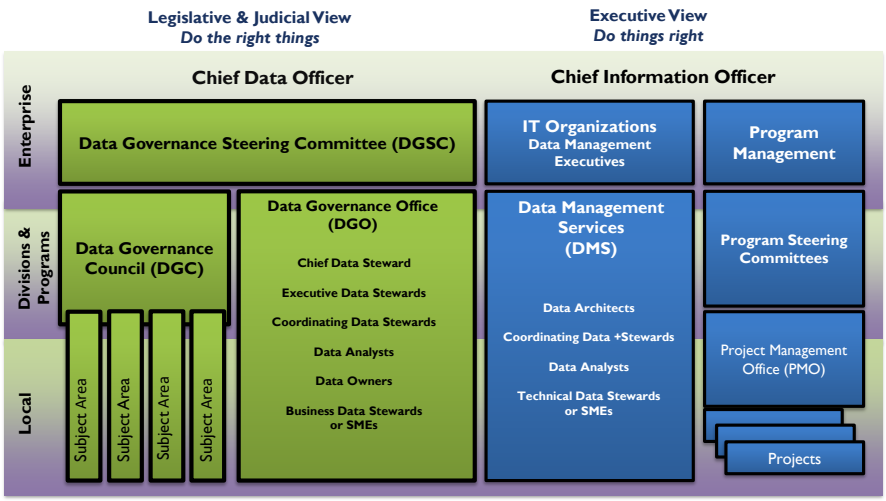
DMA DATA MANAGEMENT ADVISORS

Example Governance (RACI)

DATA SUBJECT AREA *(e.g. from CDM)*	RESPONSIBLE (R) (DOES)	ACCOUNTABLE (A) (OWNS)	CONSULTED (C)	INFORMED (I)
FINANCE	Geoff Hurst, Data Steward (Finance) Joe Smith, Data Steward (Finance) Joe Cole, Data Steward (Finance) Nobby Stiles, Data Custodian (Finance)	Gordon Banks, Data Owner (Finance)	Alan Ball, Lead Steward (Sales) Alan Knott, Data Owner (Sales)	Bobby Moore, Steward (Marketing)
SALES	Alan Ball, Lead Steward (Sales) Alan Border, Data Steward (Sales) Alan Hunt, Data Steward (Sales)	Alan Knott, Data Owner (Sales)	Gordon Banks, Data Owner (Finance) Alf Ramsey, Data Owner (Marketing)	Bobby Moore, Steward (Marketing)
MARKETING	Bobby Moore, Data Steward (Marketing) David Beckham, Data Steward (Marketing) Ben Stokes, Data Steward (Marketing)	Alf Ramsey, Data Owner (Marketing)	Geoff Hurst, Data Steward (Finance) Alan Ball, Lead Steward (Sales)	Ussain Bolt, Steward (Logistics)
ETC				

DMA DATA MANAGEMENT ADVISORS

Data Management Organizations (DMBoK)



Data Governance includes:

- legislative-like functions (defining policies, standards, and the Enterprise Data Architecture),
- judicial-like functions (issue management and escalation), and
- executive functions (protecting and serving, administrative responsibilities).

To manage risk, organizations need to adopt a representative form of data governance, so that all stakeholders can be heard.

In creating Data Management ***Services***, IT involves the Data Governance Council to:

- estimate the enterprise needs for these services,
- provide the justification for staffing and,
- provide the justification for funding

to provide these services

DMA DATA MANAGEMENT ADVISORS

Data Management Organizations (DMBoK)

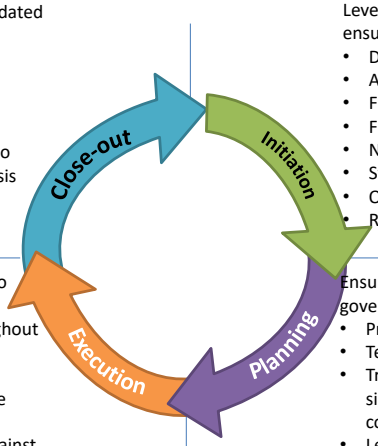
Data Governance Steering Committee	Data Governance Council (DGC)	Local Data Governance Committee(s)	Data Stewardship Teams	Data Governance Office (DGO)
The primary and highest authority organization for data governance in an organization.	Manages data governance initiatives.	Large organizations may have divisional or departmental data governance councils.	Communities of interest focused on one or more specific subject-areas or projects.	Ongoing focus on enterprise-level data definitions and data management standards.
Responsible for oversight, support, and funding of data governance activities. Facilitators responsible for council participation, communication, etc.	Manages development of policies or metrics, issues, and escalations.	Local DGC's work under the backing of an Enterprise DGC.	One or more temporary or permanent focused groups of business data stewards collaborating and consulting with project teams on data definitions and data management standards related to the area of focus.	A staff organization in larger enterprises supporting the efforts of the Data Governance Council, Data Stewardship Steering Committees, and Data Stewardship Teams
Consists of a cross-functional group of senior executives.	Consists of executives according to the operating model used.	Not recommended for smaller organizations.	Consists of business and technical data stewards and data analysts.	Consists of coordinating roles entitled data stewards or custodians, and data owners.
Typically releases funding for data governance and data governance-sponsored activities as recommended by the DGC and CDO.			Led by a coordinating data steward.	Analogous to the PMO (but for data).
This committee may have oversight from higher-level funding or initiative-based steering committees.			Typically, within an assigned subject area	DGO is across ALL Knowledge Areas. I.e. not only DG.

DMA
DATA MANAGEMENT ADVISORS

Data Governance Office

DATA GOVERNANCE HAS TOUCH POINTS THROUGHOUT THE PROJECT LIFECYCLE VIA LIAISON WITH THE DATA GOVERNANCE OFFICE (DGO, SIMILAR TO PMO)

Update Project artifacts at project close out. Updated artifacts should include:
- Updated metadata repository (Business Data Glossary & Technical Data Dictionary)
- Metrics Library
- Data / System Owners and Stewards
- Policies and Guidelines for when users need to take data offline for custom reporting / analysis
- Regulatory reporting
- "Lessons learned" report



Leverage Data Governance at project initiation to ensure consistent understanding of:
- Data needs
- Authoritative sources (using Business Glossary)
- Functional usage of data
- Frequency of need
- Needs / use beyond requester
- Stakeholders matrix
- Quality criteria
- Regulatory reporting

Throughout project leverage Data Governance to ensure:
- Appropriate stakeholder participation throughout the project
- Proper vetting of requirements and design documents to ensure data integrity issues are addressed
- Continuous analysis of new metrics / data against existing metrics & data
- Consideration of other business areas usage of this data concept

Ensure adequate resources / planning for data governance related tasks; e.g.
- Profile to assess integrity of data
- Testing using data quality scenarios
- Training and Communications plan which indicate significance of using the firm's direction for consistent information
- Leverage existing data models

DMA
DATA MANAGEMENT ADVISORS

Everyone in the Data Management Community is responsible for communicating and promoting awareness on the value of Data Governance across the organization

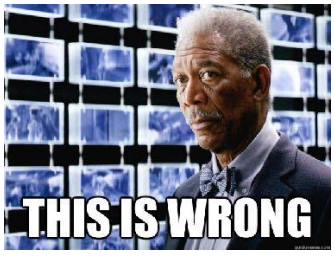
DMA
DATA MANAGEMENT ADVISORS

Guiding principles

- Data management is a shared responsibility
- Data Stewards have responsibilities in all 10 (DMBOK 1) management functions (11 in DMBOK 2)
- Every data governance / data stewardship programme is unique
- The best data stewards are found not made
- Shared decision making is the hallmark of data governance
- DG Councils/Data Stewards (legislative) while DMSG (executive)
- Data Governance occurs at enterprise and local levels
- No substitute for visionary and active IT leadership
- Centralised organisation for DM professionals is essential
- Define a formal charter for the Data Governance Council
- Data Strategy should be driven by the Business Strategy

DMA
DATA MANAGEMENT ADVISORS

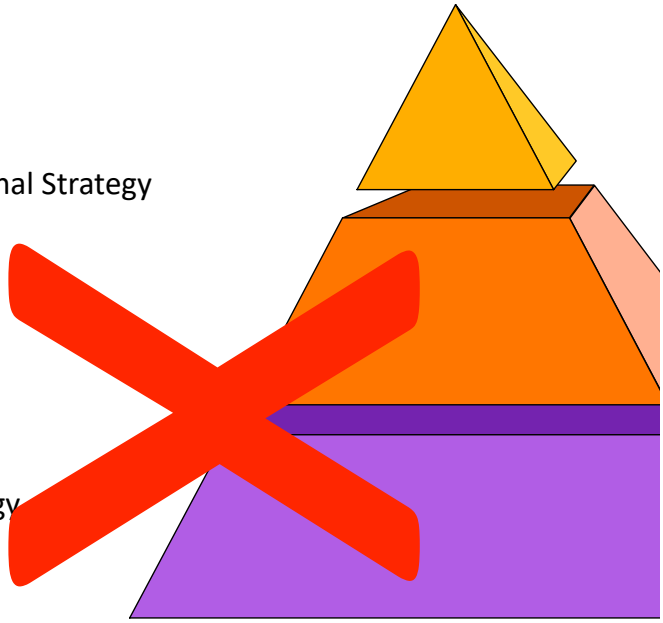
Data Strategy in Context



Organizational Strategy

IT Strategy

Data Strategy



DMA
DATA MANAGEMENT ADVISORS

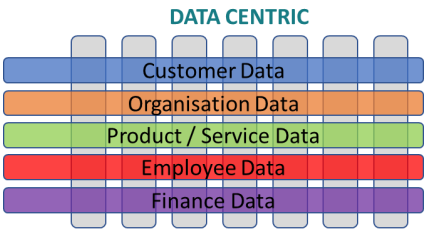
This is correct …

DMA
DATA MANAGEMENT ADVISORS

Organisational Models For DG

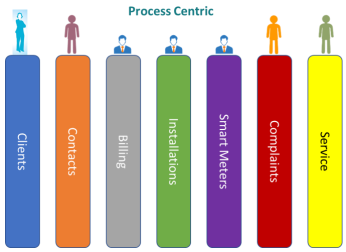
DATA CENTRIC

Business appointed full time or part time roles accountable for improvement of key data domains wherever created or used across an organisation, e.g., Data Stewardship.



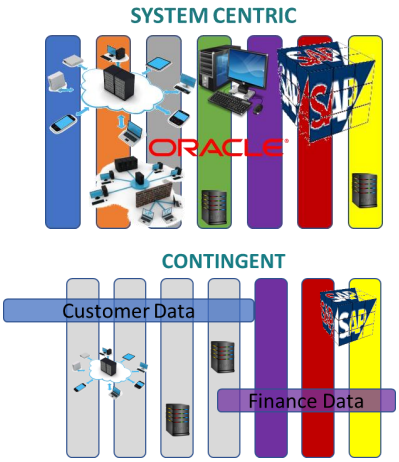
PROCESS CENTRIC

Process owners become the data owner for all data created, amended & deleted by the business process for which they are responsible.



DMA
DATA MANAGEMENT ADVISORS

Organisational Models For DG



SYSTEMS CENTRIC

System owners become the data owner for all data created, amended & deleted by the system for which they are responsible.

CONTINGENT / HYBRID

There is no single best model for data governance, either when initiating data improvement activities, or as Business As Usual.

*The **'best'** model is dependent on the type of data and the circumstances of each initiative, at each stage of maturity*

DMA
DATA MANAGEMENT ADVISORS

Operating Models For DG

DGSC reports directly to executive
Easier decision making
Easier to manage by data type
One place to stop and shop

Centralized

Decentralized (Replicated)

The main advantage of a decentralised data governance model over a centralised model is an increased level of ownership from local decision making groups.

Centralised structure, formal Data Governance Lead role serving as a single point of contact and accountability
Working groups with broad membership for facilitating collaboration and consensus building
Potentially an easier model to implement initially and sustain over time
Pushes down decision making

Hybrid

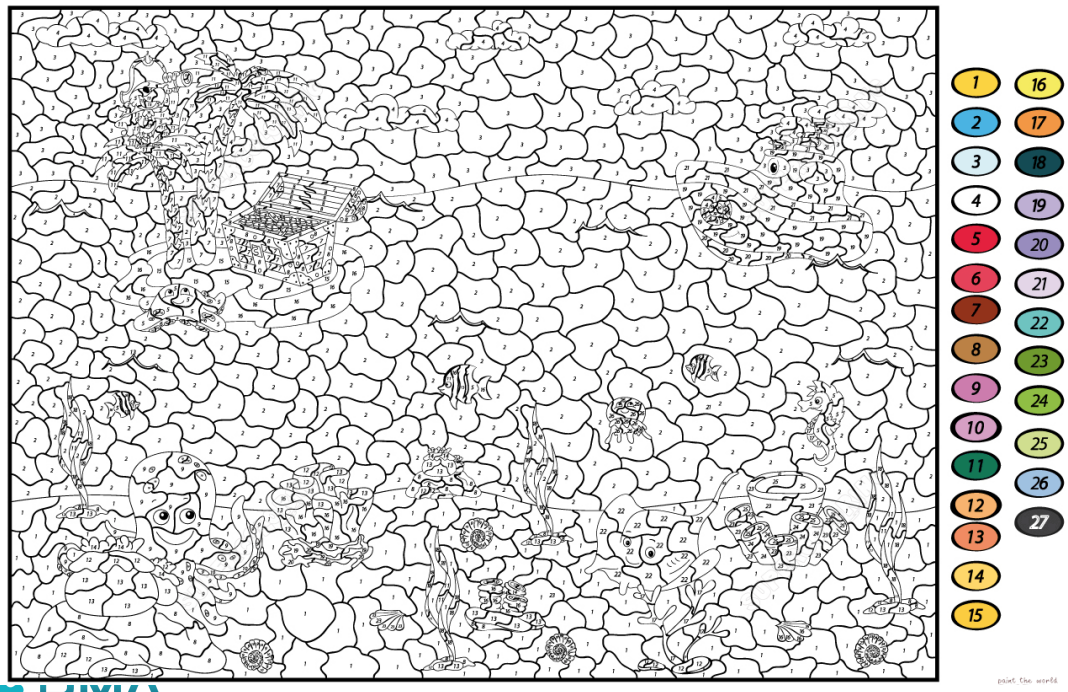
Federated

Centralized Enterprise strategy with decentralized execution and implementation
Enterprise Data Governance Lead role serving as a single point of contact and accountability
"Federated" Data Governance practices per LOB to empower divisions with differing requirements
Ability to focus on specific data entities, divisional challenges or regional priorities
Potentially an easier model to implement initially and sustain over time
Issues resolution without pulling in the whole team
Pushes down decision making

Self-organizing (Networked)

Interests of the whole group shared
No formal central structure
Genuine beneficial communities form
Natural / de facto "go to" SME's evolve
Great way to start by "stealth"
Informality of the Decentralised model made slightly more formal through a documented series of connections and accountabilities

DMA
DATA MANAGEMENT ADVISORS



DMA
DATA MANAGEMENT ADVISORS

Start Small – DG *will* ripple out

DMA

"Data Definitions for our KYC(CDE) items"

What are the top 30 CDE data items to be defined? → *High level model(s)*

DG mentoring & training ← ***How do Data Definition owners undertake their role?***

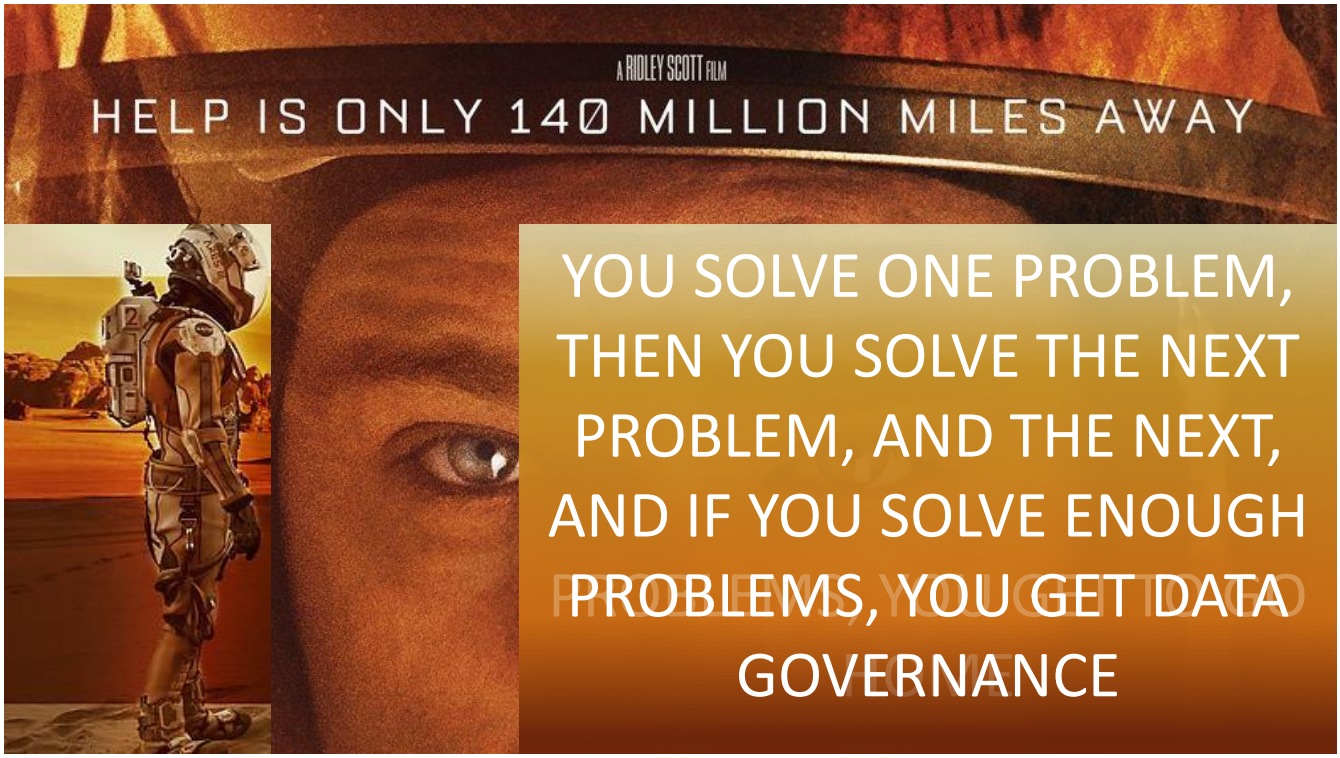
Who else needs to be Consulted, Informed about these items? → *RACI*

Map logical to physical & publish in Data Glossary ← ***What IT systems do these items exist in?***

What are the current data problems / Who decides if the definitions are conflicting? → *Ownership / Use cases*

Populate & publish Data Glossary ← ***How can we make the definitions useful & visible?***

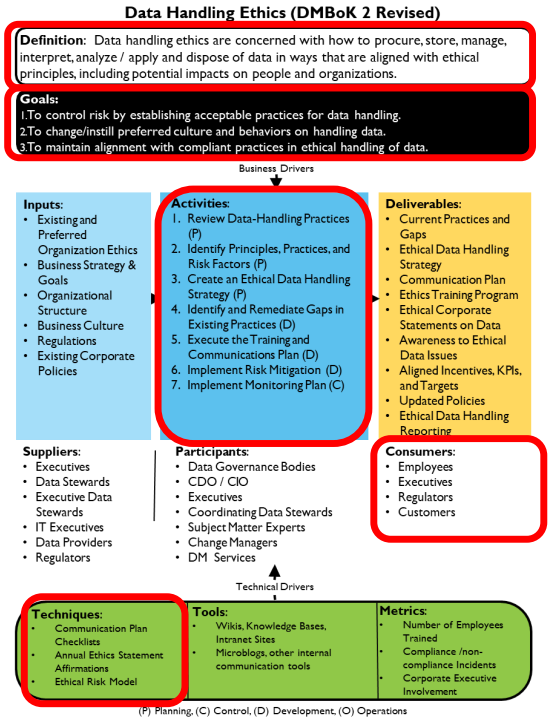
What are the quality criteria for these items? → *DQ / Ownership*



Data Governance Summary

- Data Governance is a **continuous process** of data improvement
- Data Governance is a **Business change programme**
- **IT** is a key stakeholder in DG, but **is NOT responsible for DG**
- There are **different organization models** for DG from Centralized, through Hybrid, to Federated & Self Organising Teams &
- DG is the exercise of authority and control over the management of **data assets**
- The 3 most important CSF's for Data Governance success are **Communication, Communication, Communication**
- **Everyone** in the Data Management Community is responsible for communicating and promoting awareness on the value of Data Governance across the organization
- The best **Data Stewards are found**, not made

DMA
DATA MANAGEMENT ADVISORS



Data Ethics

Data Handling Ethics are based on medical / bio ethics*

Respect for

1.autonomy,
2.beneficence,
3.non-maleficence, and
4.justice

… referred to as the four pillars of medical ethics

* The Belmont Principles were developed by the U.S. HHS in 1979 to guide the ethics of medical research, and they are also applicable as guiding principles within the field of Data Management

European Commission Article 29 Data Protection Working Party criteria to evaluate anonymization methods.

*Anonymity is protected when it is **only** possible to analyse sizeable "clusters" of individuals who **cannot** be distinguished from one another based on their attributes.*

The DMBoK2 has 3 !

The Beneficence Principle has two elements:

1. Do no harm
2. Maximize possible benefits and minimize possible harms

The Justice Principle considers the fair and equitable treatment of people.

The Respect for Persons Principle reflects the fundamental ethical requirement that people be treated in a way that respects their dignity and autonomy as human individuals.

Data Ethics

Data Handling Ethics are seeking to:

- avoid loss of reputation for the organization &
- loss of customers.

Misleading Visualisations
e.g. Leaving data points out, comparing two facts without clarifying their relationship, or ignoring accepted visual conventions, changing scale to make a trend line look better or worse.

The 5 "C's" of Data handling ethics ..

- Consent,
- Clarity,
- Consistency,
- Control *(and transparency),*
- Consequences *(and harm)*

Data Ethics

- Core **concepts** of data handling ethics are:
  1. Impact on people,
  2. the potential for misuse, &
  3. the economic value of data
- Core **deliverables** for introducing an **ethical data handling capability:**
  1. Current Practices and Gaps,
  2. Ethical Data Handling Strategy ,
  3. Communication Plan ,
  4. Ethics Training Program ,
  5. Ethical Corporate Statements on Data
- **Social licence** is the alignment between:
  1. stakeholder expectations and
  2. the organisation
- **Consent** is considered to be given when an organisation has: obtained explicit permission from an individual for the collection, use or disclosure of personal information.
- **Principles** to understand and evaluate **ethical use of data** within an organization should be based upon:
  1. fairness,
  2. respect,
  3. responsibility,
  4. integrity,
  5. quality,
  6. reliability,
  7. transparency and
  8. trust
- **Ethical responsibility** for managing data to:
  1. reduce risks of misrepresentation,
  2. misuse or
  3. misunderstanding

  lies with Data Management professionals to:
  4. manage data and
  5. manage the associated risks

Ref	Question	A	B	C	D	E
DG1	According to the DAMA DMBoK, the Data Governance steering committee (aka Data Governance Council (DGC)) is the primary and highest authority organization for data governance in an organization. Who should typically chair this Council?	The Chief Information Officer (CIO)	Chief Data Steward (Business) / Chief Data Officer	The chair should rotate across the Data Owners	The Chief Data Architect	Any Executive / C-level participant in the DGC
DG2	What are the primary characteristics of a data steward?	A business role appointed to take responsibility for the quality and use of their organization's data assets.	Analyzing data quality	The manager responsible for writing policies and standards that define the data management program for an organization.	Identifying data problems & issues	The data analyst who is the subject matter expert (SME) on a set of reference data.
DG3	Which of these is NOT true of Data Governance?	DG is a continuous process of data improvement	IT is a key stakeholder in DG	A DG initiative should always be led by the IT department	There are different organization models for DG	DG is the exercise of authority and control over the management of data assets
DG4	Who is responsible for communicating and promoting awareness on the value of Data Governance in the organisation?	Central Communications and Corporate Awareness	Data Stewards	The Chief Executive Officer	Senior Management Executive Forum	Everyone in the Data Management Community
DG5	Communicating the value of Data Governance can be approached in a number of ways. Which of the following approaches is NOT a recognised way of doing this?	Providing only negative communications on ongoing data issues to key executive stakeholders	Maintaining an intranet website	Publishing a regular newsletter via hardcopy or email	Promote participation in a DM forum or community	Creating a series of "elevator pitches" for the appropriate audience
DG6	When considering a Data Governance program, communication is a key element. There are many ways of managing this communication, with one of the most effective being a Data Management intranet. Which of the following would you typically NOT put onto such an communication vehicle	Description of the DG organisation, it's key members and contact details	Executive message regarding significant data management issues	The Data Steward team profiles	Raw data results of an investigation into a possible data privacy breach	Link to a "raise an issue" log
DG7	How can the Data Governance process in an organization best support the requirements of various regulatory reporting needs?	by providing a business glossary based look up facility for data definitions	by ensuring that data is properly categorized, owned, understood, defined, documented, and controlled	by performing an as-is data audit	by highlighting the challenges of multiple data definitions within the enterprise	by creating a map of where the enterprise data is located in IT systems
DE1	One way of defining ethics is:	"doing it wrong, and then apologising".	"doing it right when someone is looking".	"doing it wrong, and then expertly covering it up".	"doing it wrong, and failing to covering it up".	"doing it right when no one is looking".
DE2	The ethics of data handling, centre on several core concepts. They are:	impact on people, potential for misuse and economic value.	impact on people, potential for re-use and storage cost.	access to data, potential for misuse and storage cost.	accurate business glossary, data quality and reference data.	privacy, security and authorisation.
DE3	In data handling ethics, 'social licence' refers to the alignment between:	stakeholder demands and technology deliverables.	stakeholder expectations and the organisation	public perception and published fact.	society's needs and their right to access data.	social and political decision matrices.

AFTER QUIZ 11

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)
8. Document, Records & Content Mgt (3)
9. Architecture & lifecycle (4)
10. Metadata management (6)
11. Data Governance & Ethics (10)

Maximum possible score = 72

60% (CDMP Associate) = 44

70% (CDMP Practitioner) = 51

80% (CDMP Master) = 58

Data Integration & Interoperability

Data Management Process	2%
Big Data	2%
Data Architecture & Lifecycle Management	6%
Document, Records & Content Management	4-5%
Data Ethics	2%
Data Governance	11%
Data Integration & Interoperability	6-7%
Master & Reference Data Management	11%
Data Modelling & Design	11%
Data Quality	11%
Data Security	6%
Data Storage & Operations	4-5%
Data Warehousing & Business Intelligence	11%
Metadata Management	11%

DMA
DATA MANAGEMENT ADVISORS

Data Integration and Interoperability (DMBoK 2 revised)

Definition: Data Integration is the movement and consolidation of data within and between data stores, applications and organizations. Data Interoperability is the ability for multiple systems to communicate.

Goals:
1. Make data available from disparate sources in the format and timeframe needed.
2. Lower cost and complexity of managing solutions by developing shared models and interfaces.
3. Identify meaningful events and automatically trigger alerts and actions.
4. Support business intelligence, analytics, master data management, and operational efficiency efforts.

Business Drivers

Inputs:
- Business Goals & Strategies
- Data Needs & Standards
- Regulatory, Compliance, & Security Requirements
- Data, Process, Application, and Technical Architectures
- Data Semantics
- Source Data

Activities:
1. **Plan & Analyze (P)**
   1. Define data integration and lifecycle requirements
   2. Perform Data Discovery
   3. Document Data Lineage
   4. Profile Data
   5. Collect Business Rules
2. **Design DII Solutions (P)**
   1. Design DII Architecture
   2. Model Data Hubs, Interfaces, Messages, and Data Services
   3. Map Data Sources to Targets
   4. Design Data Orchestration
3. **Develop DII Solutions (D)**
   1. Develop Data Services
   2. Develop Data Flows
   3. Develop Data Migration Approach
   4. Develop a Publication Approach
   5. Develop Complex Event Processing Flows
4. **Implement and Monitor (O)**
   1. Maintain DII Metadata

Deliverables:
- DII Architecture
- Data Exchange Specifications
- Data Access Agreements
- Data Services
- Complex Event Processing Thresholds and Alerts

Suppliers:
- Data Producers
- Architecture Board
- IT Steering Committee
- Executives and Managers
- Subject Matter Experts

Participants:
- Data Architects
- Business and Data Analysts
- Data Modelers
- DII Developers
- Project and Program Managers
- Data Stewards

Consumers:
- Customers, Partners
- Information Consumers
- Knowledge Workers
- Data Scientist

Technical Drivers

Techniques:
- Loosely coupled applications
- Minimize the number of interfaces
- Create standard canonical interfaces
- Service Orchestration

Tools:
- Data Transformation Engine
- Business Rules Engine
- Data Virtualization Server
- Enterprise Service Bus
- Data and Process Modeling Tools
- Data Profiling Tool
- Metadata Repository

Metrics:
- Data Availability
- Data Volumes and Speed
- Solution Costs and Complexity

(P) Planning, (C) Control, (D) Development, (O) Operations

DMA DATA MANAGEMENT ADVISORS

Goals

Applying Data I&I practices and solutions aims to:

- Make data available in the format and timeframe needed by data consumers, both human and system
- Consolidate data physically and virtually into data hubs
- Lower cost and complexity of managing solutions by developing shared models and interfaces
- Identify meaningful events (opportunities and threats) and automatically trigger alerts and actions
- Support business intelligence, analytics, master data management, and operational efficiency efforts

DMA
DATA MANAGEMENT ADVISORS

Core Concepts:

Integration: Movement and consolidation of data into consistent forms (physical or virtual) within and between data stores, applications, and organizations.

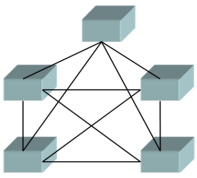
Interoperability: The ability for multiple systems to communicate through data that does not require processing on either side

Common use cases:

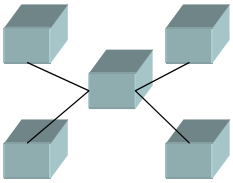
1. Integration of data between data stores, including ingesting from and disbursing data to external sources
2. Consolidation of data stores, including application consolidation, data hub management, and mergers and acquisitions
3. Distribution across data stores and data centres
4. Moving data into archives, and updating data from one archive technology to another to enable future use

DMA
DATA MANAGEMENT ADVISORS

Point to Point

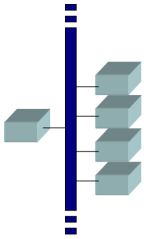


Hub Distribution

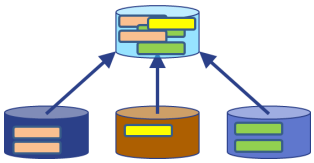


Message Synchronisation & Propagation

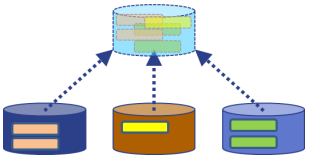
Bus Distribution



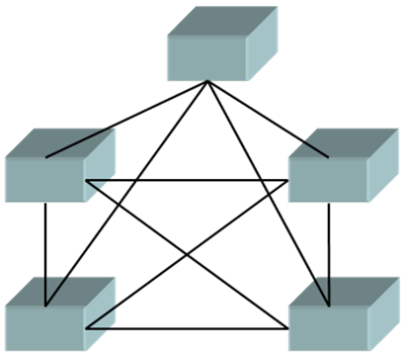
ETL, ELT & CDC



Abstraction / Virtual Consolidation



Point to Point

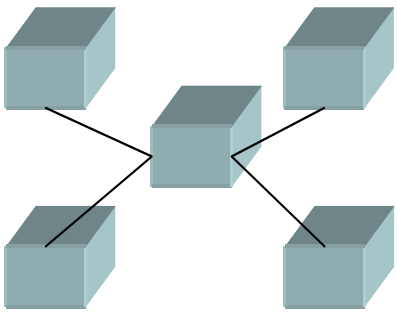


- Good for a very limited number of interfaces
- Short life expectancy of interfaces
- Need for speed

Issues:

- Approach does not easily / predictably scale.
- Run-time: performance issues may arise when many systems want to get the same data from the same source.
- Design-time: issues arise when multiple systems require different versions or formats of the data.
  - The source systems support teams must build or support separate interfaces for each specific use.
- Support: While point-to-point interfaces may be most performant, the whole ecosystem may suffer because exterior effects are not considered.

Hub Distribution



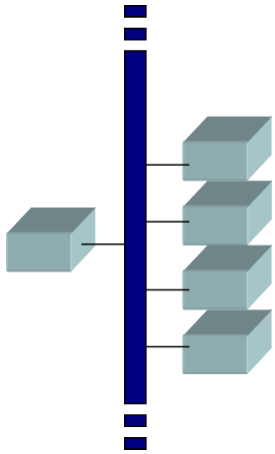
- Data updates (published) into a central hub repository which distribute them to all authorized (subscribing) applications and may also keep a copy in a repository.
- An architecture formerly used by Master & Reference Data Management systems (MDM); if the messaging hub is modified to persist data, the hub can act as the Record of Reference.
- Messaging hub / repository interprets each transaction, knows how to transform it, and which applications are authorized to receive the updates.

Potential Issues:

- Situations (e.g., security or regulations) where data should not be shared in a hub. Use a small dedicated secured hub for that data.
- Latency may be unacceptable, or performance is insufficient.
- Single point of failure

DMA
DATA MANAGEMENT ADVISORS

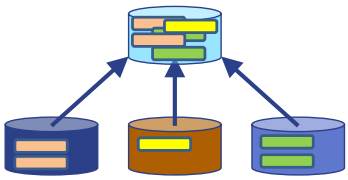
Bus Distribution



- Bus Distribution model “pushes” data into a central service which then pushes it to authorized applications or the application can “pull” the data.
- Data is not permanently retained and therefore it cannot act as the Record of Reference.
- Often described as publish-subscribe.
- Basis of a Services Oriented Architecture (SOA).
- Well-suited for applications using a one-to-many distribution which do not have data retention requirements for the distribution tool
- Compared to the Hub model, this model is more scalable, can offer better performance, and is well-suited for a multi-mega-centre environment.
- It has higher initial build costs but with lower maintenance costs.
- Same single point of failure concern as the Hub model.

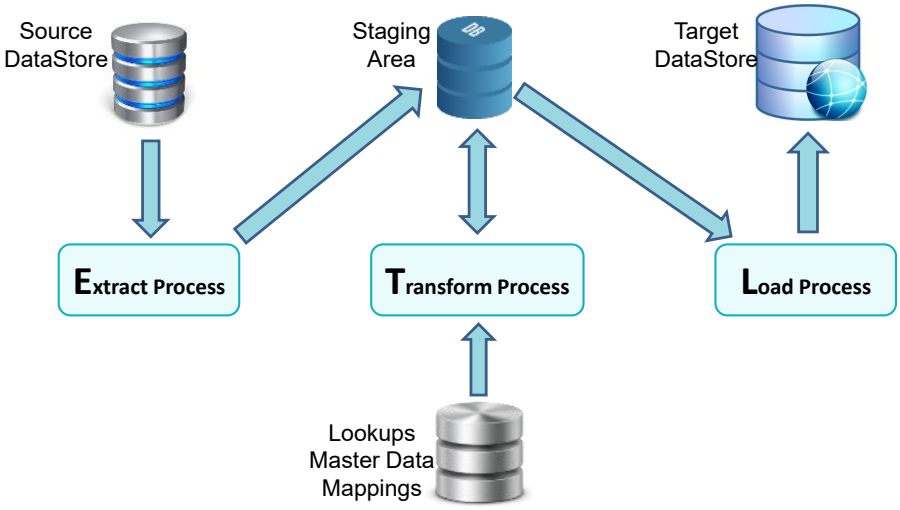
Physical movement & consolidation

ETL, ELT & CDC

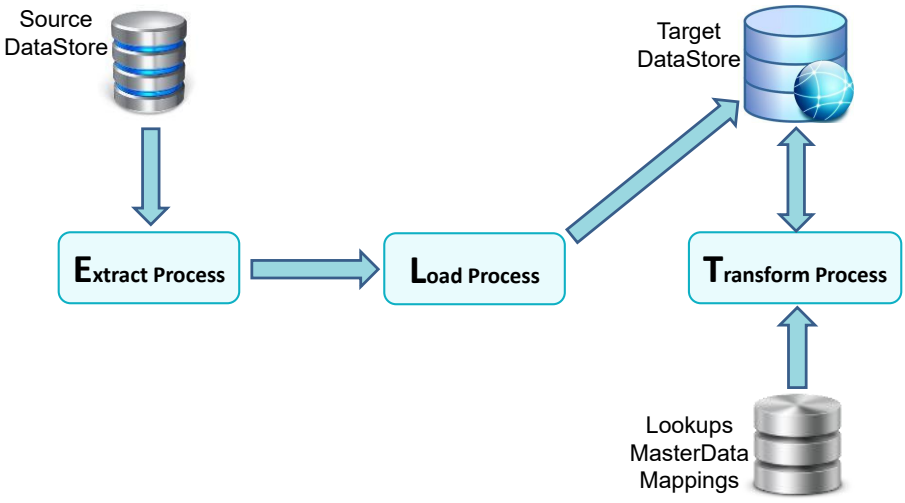


- Batch distribution for the mass movement of data collected over time from a source data structure and distributed as a batch to another.
- Extract Transform & Load (ETL) and Extract Load and Transform (ELT tools facilitate this process with scheduling, parallel processing and complex data transformation, cross reference and data mapping capabilities.
- Batch may contain incremental or full downloads of the data. This is probably the most commonly used approach to Data Integration.
- ETL / ELT is typically run according to a schedule (e.g. the “overnight load of a Data Warehouse) and is used for bulk data movement, usually in in batch.
- Change Data Capture (CDC) is event driven (e.g. the stock level has fallen below critical) and delivers real-time incremental replication.

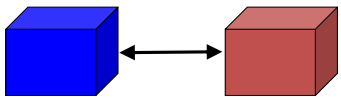
ETL Process Flow



ELT Process Flow



Message Synchronisation & Propagation



- Message based synchronisation and data propagation is used for application to application integration.

2 main genres:

- Enterprise Application Integration (EAI)
- Enterprise Service Bus (ESB)
- Can be based on a Hub or a Bus.
- Both of these are used primarily for the purpose of event driven business process automation.
- May be loosely or tightly coupled

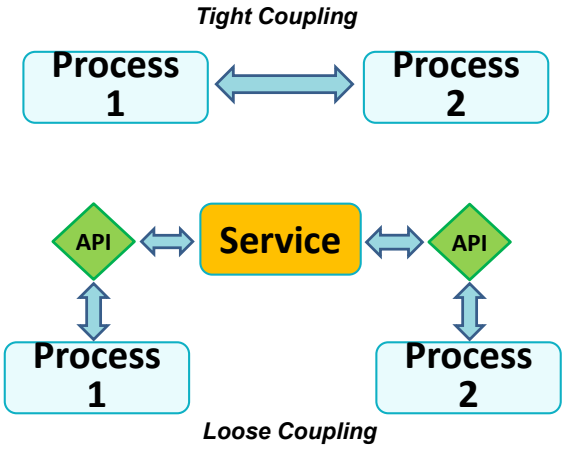
DMA DATA MANAGEMENT ADVISORS

Application Coupling

Coupling describes how an interface connects any two systems.

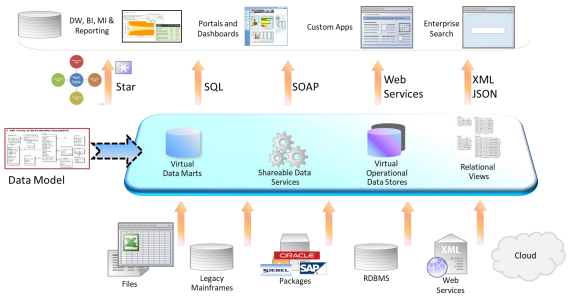
Tight coupling means that the sources and targets are pre-defined and directly named.

Loose coupling allows the sources and targets to **not be linked directly;** instead, a service is used to allow sources and targets to remain anonymous to each other, and instead, the applications use APIs to link to the service.



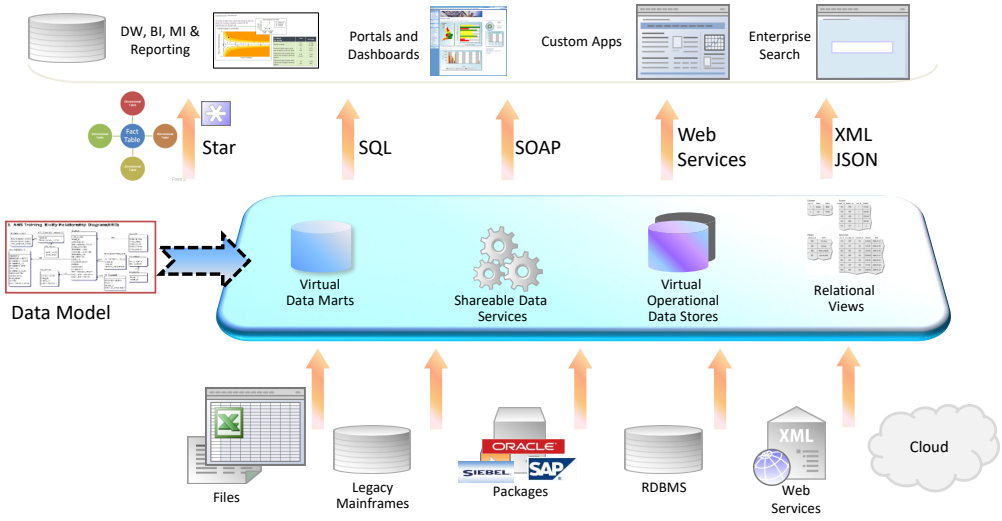
DMA DATA MANAGEMENT ADVISORS

Abstraction / Virtual Consolidation



- Virtual distribution leaves data in situ while building an intelligent layer on top of it that can locate and transform data on the fly.
- Rapidly emerging option for organizations with large legacies that they intend to retain without making major investments in transforming them.
- Consuming applications see data as though it's in their own systems
- Often referred to as Enterprise Information Integration (EII)
- No need to touch underlying source data
- Very rapid time to solution
- Easily mix structured & unstructured data sources
- Present data to applications as SOAP, SQL, STAR, ODBC, XML, JSON …..

DMA DATA MANAGEMENT ADVISORS



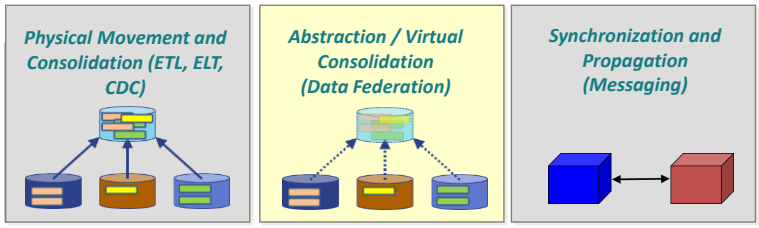
DMA DATA MANAGEMENT ADVISORS

TIBCO® Data Virtualization

denodo DATA VIRTUALIZATION

DMA DATA MANAGEMENT ADVISORS

Key Differences with Data Virtualisation



Approach	ETL/ELT	CDC	*Data Virtualization*	EAI / ESB / API
Purpose	DB ⇔ DB	DB ⇔ DB	*DB ⇔ Application*	Application ⇔ Application
Attribute	Scheduled	Event Driven	*On Demand*	Event Driven

DMA DATA MANAGEMENT ADVISORS

Other DII Terms & Considerations

Latency: The **time difference** between when **data is generated** in the source system and when the **data is available** for use in the target system. Different approaches to data processing result in different degrees of data latency. Latency can be high (batch) or low (event-driven) to very low (real-time synchronous).

Asynchronous Data Flow: This allows the source system to **continue processing without waiting** for the receiving system's acknowledgment. It implies that **systems can operate independently**, and updates are near-real-time, usually seconds or minutes apart. This approach does not disrupt the source application's processing if target applications are unavailable. However, it results in a delay between updates made in the source and those relayed to target data sets.

Synchronous integration: This requires an **executing process to wait for confirmation** from other applications before proceeding. It **ensures data synchronization** but can process fewer transactions due to waiting times. If any application required for the update is unavailable, the transaction cannot be completed. This method keeps data sets perfectly in sync, often using database capabilities like two-phase commits, especially in financial institutions to synchronize transaction and balance tables.

Ref	Question	A	B	C	D	E
DII 1	The need to manage data movement efficiently is a primary driver for:	Data Integration and Interoperability	Data Storage and Operations	Data Warehousing and Business Intelligence	Document and Content Management	Data Security
DII 2	The acronym ETL most commonly stands for:	Export Transform Log	Extract Transform Load	Extend Trim Load	Efficient Trace Logging	Extract Transpose Leverage
DII 3	Mapping requirements and rules for moving data from source to target enables:	load	extract	transformation	analysis	backups
DII 4	When integrating two data stores using batch or real-time synchronous approaches, results in a difference in:	data quality	lethargy	source of truth	timestamping	latency
DII 5	If two data stores are able to be inconsistent during normal operations, then the integration approach is:	Streaming	Synchronous	Faulty	Asynchronous	Uncontrolled
DII 6	A 'Content Distribution Network' supporting a multi-national website is likely to use:	a replication solution	an extract transform and load solution	a database backup and restore solution	an archiving solution	a records disposal solution
DII 7	Data that is used infrequently or not at all may be moved to an alternative data store. This is called:	replication	analysis	archiving	auditing	authentication
DII 8	Three common interaction models for data integration are:	point to point, hub and spoke, publish and subscribe.	point to point, wheel and spoke, public and share.	plane to point, harvest and seed, publish and subscribe.	straight copy, curved copy, roundabout copy.	record and pass, copy and send, read and write.

AFTER QUIZ 12

1. General (6)
2. Data Quality (9)
3. Data Storage & Operations (3)
4. Master & Reference Data (9)
5. DW / BI + Big Data (7)
6. Data Modelling (9)
7. Data Security (6)
8. Document, Records & Content Mgt (3)
9. Architecture & lifecycle (4)
10. Metadata management (6)
11. Data Governance & Ethics (10)
12. Data Integration & Interoperability (8)

DMA

Maximum possible score = 80

60% (CDMP Associate) = 48

70% (CDMP Practitioner) = 56

80% (CDMP Master) = 64

DMBoK$_2$ and CDMP® Preparation

Live and On-Demand Classes

REGISTERED DAMA International EDUCATION PROVIDER

Data Management Fundamentals
https://www.dataversity.net/dmbok-and-cdmp-preparation-learning-plan/

Data Governance
https://training.dataversity.net/learning-paths/dgs0-dmbok-and-cdmp-preparation-data-governance-specialist-learning-plan

Data Modelling
https://training.dataversity.net/learning-paths/dms0-dmbok-and-cdmp-preparation-data-modeling-specialist-learning-plan

Data Quality Management
https://training.dataversity.net/learning-paths/dqs0-dmbok-and-cdmp-preparation-data-quality-specialist-learning-plan

Master & Reference Data Management
https://www.dataversity.net/

Metadata Management
https://www.dataversity.net/

- CDMP Online learning program
- Approved by DAMA-I
- Part of DATAVERSITY training center
- Based on DMBoK CDMP syllabus

20% discount code for DAMA UK
"dmadvisors"

DMA
DATA MANAGEMENT ADVISORS

Contact

info@DMAdvisors.co.uk

+44 7973 184475 (mobile) +44 1225 923000 (office)

@inforacer

uk.linkedin.com/in/christophermichaelbradley/

infomanagementlifeandpetrol.blogspot.com

TRAINING

ADVISORY

CONSULTING

CERTIFICATION

DMA
DATA MANAGEMENT ADVISORS

DMA
DATA MANAGEMENT ADVISORS