UMBC

DATA 605
Ethical & Legal Issues in Data Science

SPRING 2022

SUNELA THOMAS

MARCH 17, 2022

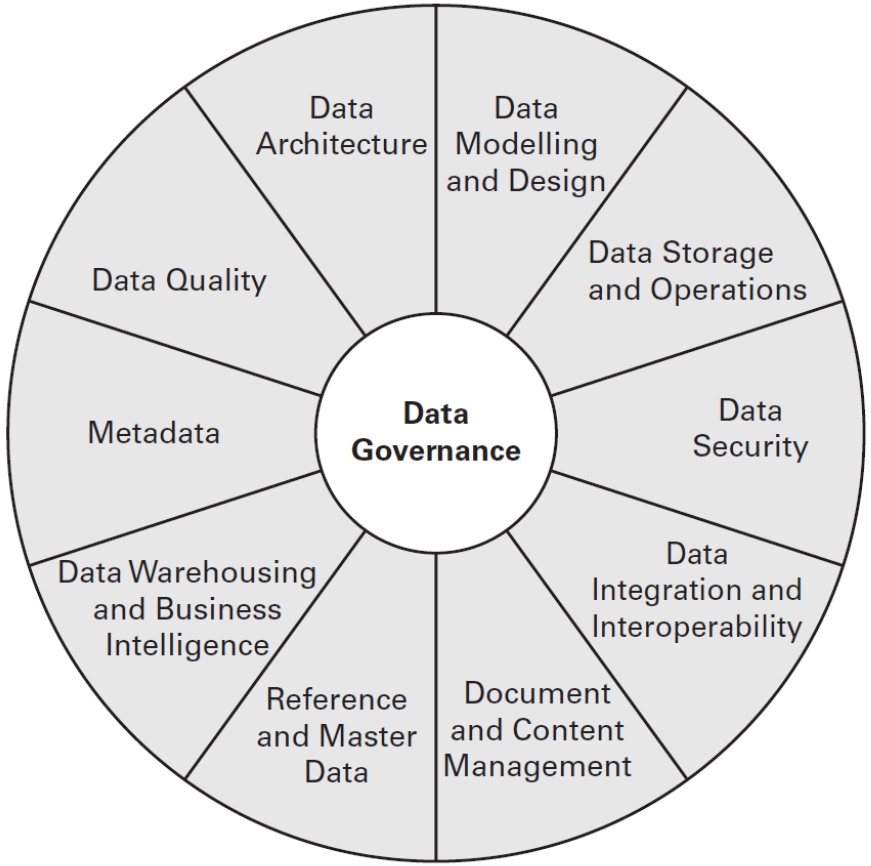
AGENDA

- Questions?
- Paper #2 posted
- Common Ethical Challenges for Data Practitioners and Users
- Breakout Discussion
- Ethics in the Data Management Body of Knowledge

Ethical Challenges in Appropriate Data Collection & Use

1. How can we properly acknowledge and respect the purpose for, and context within which, certain data was shared with us or generated for us?
2. How can we avoid unwarranted or indiscriminate data collection?
3. Have we adequately considered the ethical implications of selling or sharing subjects' data with third-parties?
4. Have we given data subjects appropriate forms of choice in data sharing?
5. Are the terms of our data policy laid out in a clear, direct, and understandable way, and made accessible to all data subjects?
6. Are data subjects given clear paths to obtaining more information or context for a data practice?
7. Are data subjects being appropriately compensated for the benefits/value of their data?
8. Have we considered what control or rights our data subjects should retain over their data?

DAMA DMBOK Wheel



SOURCE Copyright © 2017 DAMA International

Data Governance

Process of managing the availability, usability, integrity and security of the data in enterprise systems based on internal data standards and policies that also control data usage.

Well designed data governance program typically includes a governance team, a steering committee that acts as the governing body and a group of data stewards

Why Data Governance Matters

To avoid inconsistent data silos in different departments and business units

To agree on common data definition for a shared understanding of data

To improve data quality through efforts to identify and fix errors in data sets

To increase analytics accuracy and give decision-makers reliable information

To implement and enforce policies that help prevent data errors and misuse

To help ensure compliance with data privacy laws and other regulations

Data Governance Goals and Benefits

Break down data silos

Aims to harmonize data

Data is used properly

Strike a balance between data collection practices and privacy mandates

Improve data quality

Lower data management costs

Increase access to needed data for data scientists

Improve business decision making

Who is Responsible for Data Governance

Chief Data Officer

Data Governance Committee

Data Stewards

Components of a Data Governance Framework

Identify owners or custodians of the different data assets across an enterprise

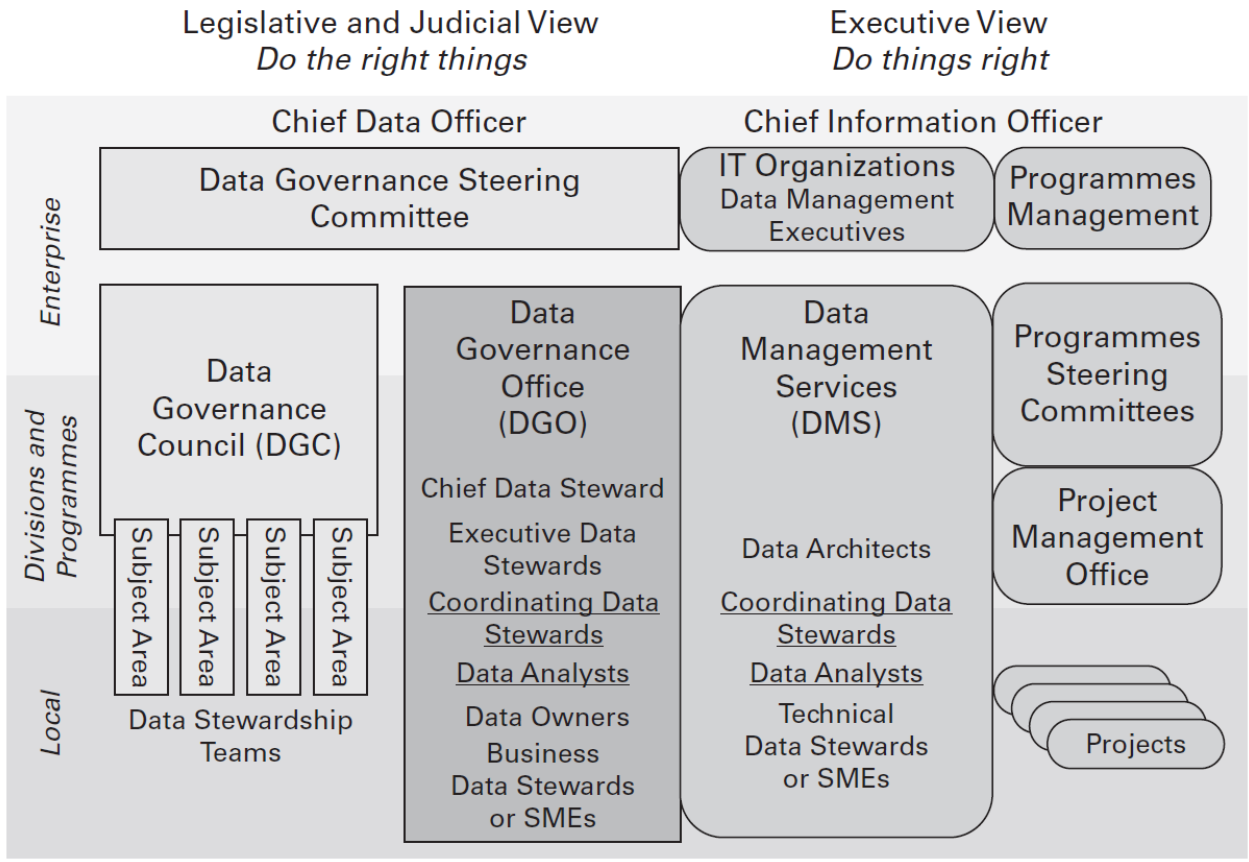
Policies and data standards

Data mapping & classification

Business glossary

Data catalog

Data Governance in the Organization



SOURCE DAMA International

Data Stewards

Chief Data Stewards

Executive Data Stewards

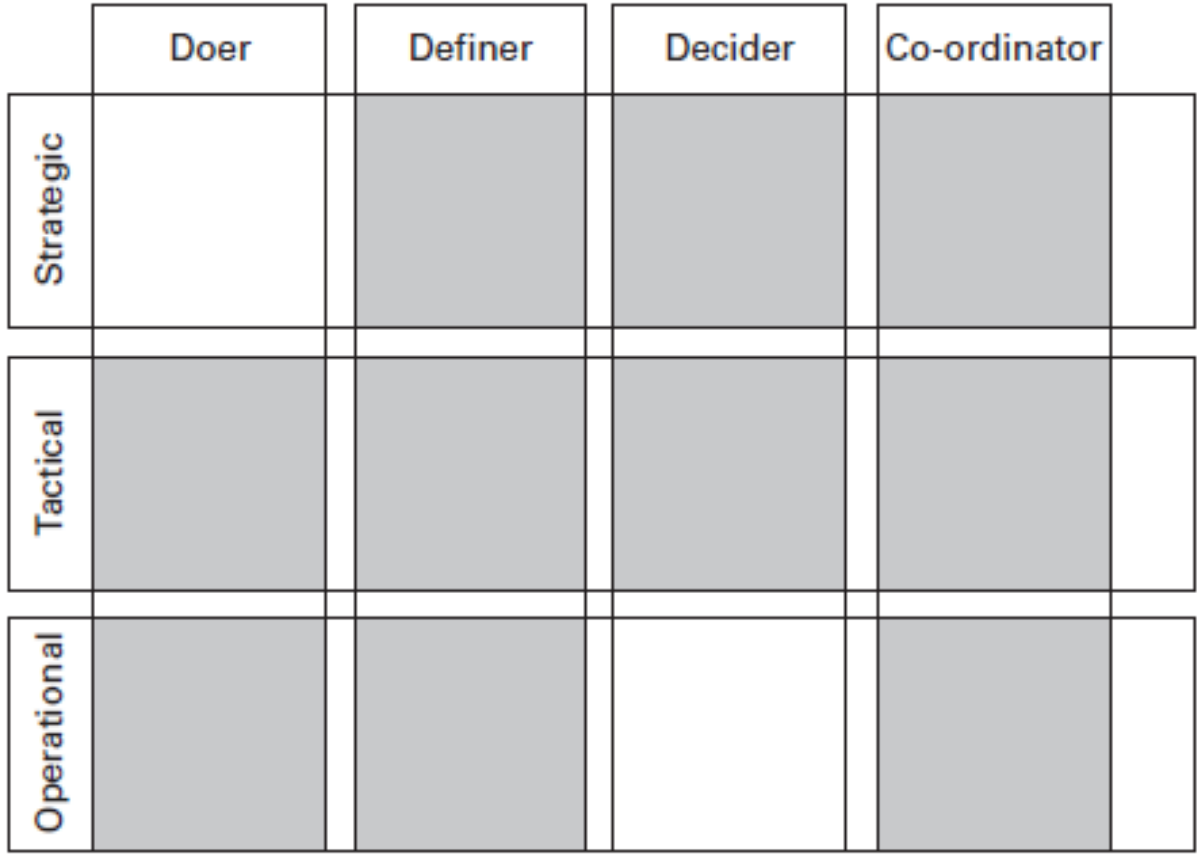
Enterprise Data Stewards

Business Data Stewards

Technical Data Stewards

Co-ordinating Data Stewards

Sample Framework for a Consulting Business



Data Architecture (Zachman framework example)

	DATA *What*	FUNCTION *How*	NETWORK *Where*	PEOPLE *Who*	TIME *When*	MOTIVATION *Why*
Objective/Scope (contextual) *Role: Planner*	List of things important in the business	List of Business Processes	List of Business Locations	List of important Organizations	List of Events	List of Business Goal & Strategies
Enterprise Model (conceptual) *Role: Owner*	Conceptual Data/ Object Model	Business Process Model	Business Logistics System	Work Flow Model	Master Schedule	Business Plan
System Model (logical) *Role:Designer*	Logical Data Model	System Architecture Model	Distributed Systems Architecture	Human Interface Architecture	Processing Structure	Business Rule Model
Technology Model (physical) *Role:Builder*	Physical Data/Class Model	Technology Design Model	Technology Architecture	Presentation Architecture	Control Structure	Rule Design
Detailed Reprentation (out of context) *Role: Programmer*	Data Definition	Program	Network Architecture	Security Architecture	Timing Definition	Rule Speculation
Functioning Enterprise *Role: User*	Usable Data	Working Function	Usable Network	Functioning Organization	Implemented Schedule	Working Strategy

Data Modelling

Conceptual Data Model

Logical Data Model

Physical Data Model

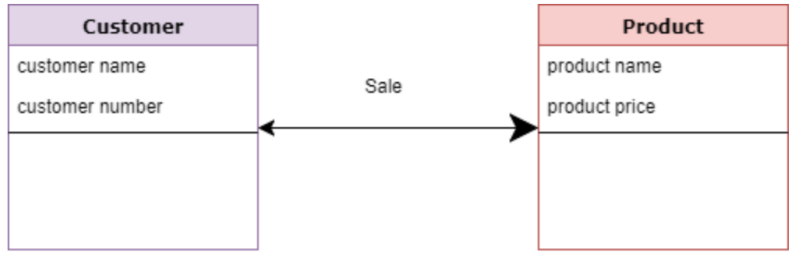
Conceptual Data Model

Organized view of database concepts and their relationships.

Purpose is to establish entities, their attributes and relationships

Data model example:

- Customer and Product are two entities. Customer number and name are attributes of the Customer entity
- Product name and price are attributes of product entity
- Sale is the relationship between the customer and product



Conceptual Data Model

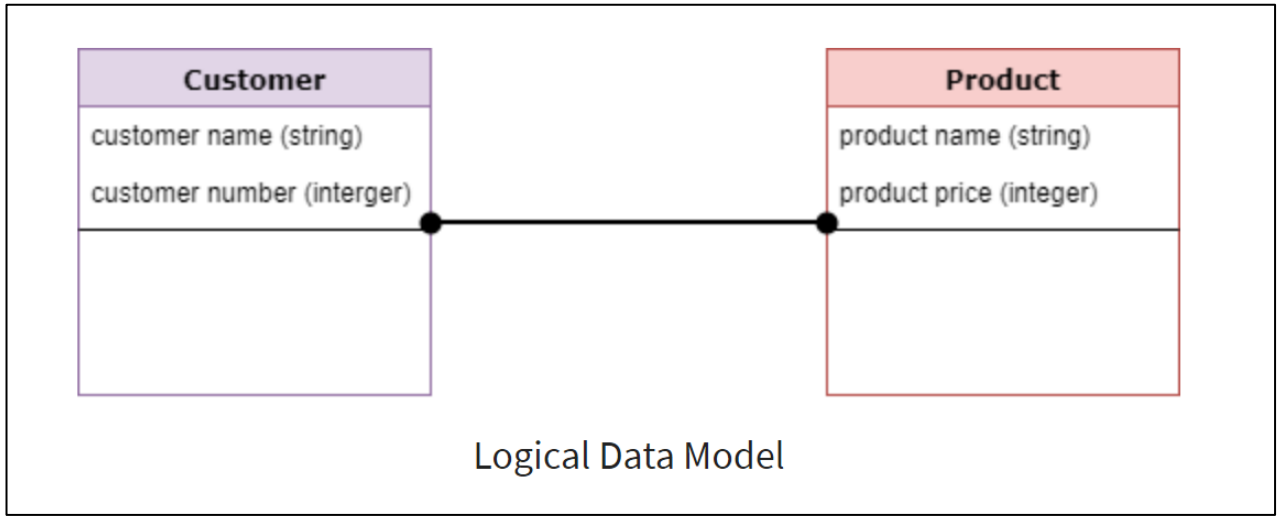
Conceptual Data Model Characteristics

- Offers Organization-wide coverage of the business concepts.
- This type of Data Models are designed and developed for a business audience.
- The conceptual model is developed independently of hardware specifications like data storage capacity, location or software specifications like DBMS vendor and technology. The focus is to represent data as a user will see it in the “real world.”

Logical Data Model

Used to define the structure of data elements and to set relationships between them

Adds further information to the conceptual data model elements



Logical Data Model

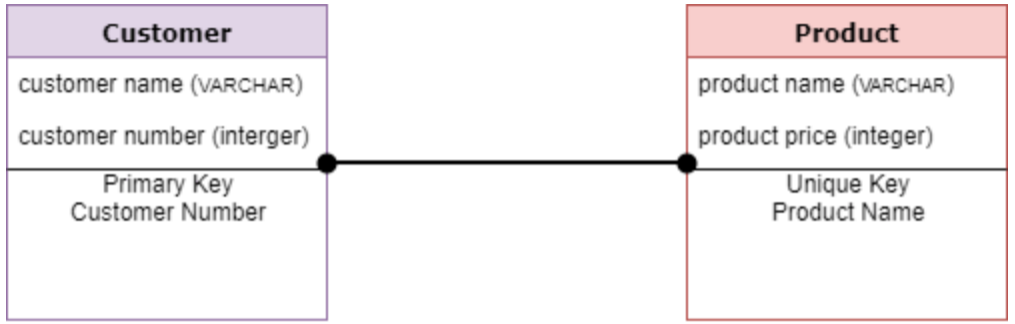
Logical Data Model Characteristics

- Describes data needs for a single project but could integrate with other logical data models based on the scope of the project.
- Designed and developed independently from the DBMS.
- Data attributes will have datatypes with exact precisions and length.
- Normalization processes to the model is applied typically till 3NF.

Physical Data Model

Describes a database-specific implementation of the data model

Offers database abstraction and helps generate schema



Physical Data Model

Physical Data Model Characteristics

- The physical data model describes data need for a single project or application though it maybe integrated with other physical data models based on project scope.
- Data Model contains relationships between tables that which addresses cardinality and nullability of the relationships.
- Developed for a specific version of a DBMS, location, data storage or technology to be used in the project.
- Columns should have exact datatypes, lengths assigned and default values.
- Primary and Foreign keys, views, indexes, access profiles, and authorizations, etc. are defined.

Advantages of Data Model

- The main goal of a designing data model is to make certain that data objects offered by the functional team are represented accurately.
- The data model should be detailed enough to be used for building the physical database.
- The information in the data model can be used for defining the relationship between tables, primary and foreign keys, and stored procedures.
- Data Model helps business to communicate the within and across organizations.
- Data model helps to documents data mappings in ETL process
- Help to recognize correct sources of data to populate the model

Disadvantages of Data Model

- To develop Data model, one should know physical data stored characteristics.
- This is a navigational system produces complex application development, management. Thus, it requires a knowledge of the biographical truth.
- Even smaller change made in structure require modification in the entire application.
- There is no set data manipulation language in DBMS.

Data Modelling and Ethics

Ethics Topic	Opposite Perspective	Dilemma	Data Modeling Application
Righteousness (do the 'right' thing)	There is no 'right' thing or do it the right way. Infinite perspectives.	- What if asked to do something that you think is wrong? - Do no harm – who defines harm or what the 'right' way is? - What if your way is the 'right' way but others disagree? - Is it better to look out for the 'greater good' or is it sometimes OK to look after oneself?	- How to be of the best service? - Is there one right way to model something? Is there a right or wrong in data modeling? Or is it a case of what is most useful in the situation? - Big data – do we need a data model? Or is it a waste of time? - A normalized model is the 'right' way for this application but others disagree. - This level of flexibility is correct! - Business has requirements, but this is correct technology wise. - The DBA wants to ignore model and provide performance.

Data Modelling and Ethics

Ethics Topic	Opposite Perspective	Dilemma	Data Modeling Application
Speak/model the truth and act on it.	What is the truth?	- What are the facts versus opinions? - How can we speak with the greatest amount of integrity? - How can we provide models with integrity? - Should we always do what we say?	- What are the real facts in the model and what is opinion/judgment/perspectives? (need the most flexible model – is that a fact?) - What if we are asked to model something that is not true? (e.g., a person can only play one role, only 2 lines for an invoice) - Model the truth? Can we only approximate and get closer to the truth? - What is the single version of the truth – who is right? - Be transparent with mistakes?

Data Modelling and Ethics

Ethics Topic	Opposite Perspective	Dilemma	Data Modeling Application
Confidentiality	Whistle Blowing	- Honor confidentiality agreement if you discover something illegal or unethical? (E.g., Snowden, Assange/WikiLeaks)	- Disclose use of data for unethical purpose? E.g., Using personal data such as social media, genome data, or private information for understanding behavior? - What if asked to model something you shouldn't? e.g., credit card information which is illegal to store?

Data Modelling and Ethics

Ethics Topic	Opposite Perspective	Dilemma	Data Modeling Application
Stealing, plagiarism, taking without permission	Re-use	When can you take and when do you need permission? When is something your own?	- What if your data model was taken from a 're-usable model' and you didn't get permission or you didn't disclose and have full transparency? - Who owns the intellectual property of a model? When is it yours? - When do you need permission to re-use models or modeling ideas?

Data Modelling and Ethics

Ethics Topic	Opposite Perspective	Dilemma	Data Modeling Application
Responsible for designing protection/privacy/ security/encryption	Is this the data modeler's job?	How much to design and what is in scope?	- What is in scope for modeling security/privacy and in which model (conceptual, logical, physical?) - What about storing data in the cloud? Less control over security/privacy?

Data Quality Management

Planning, implementation, and control activities that apply quality management techniques to data, to assure it is fit for consumption and business purposes

Ethics in Information Quality:

- Focus on the needs of their customers
- Drive out fear and encourage pride in a job well done
- Focus on continuous improvement

Data science and big data

Technologies and tools of data science and big data are:

- Machine Learning
- Sentiment Analysis
- Data and Text Mining
- Predictive Analysis
- Prescriptive Analysis
- Unstructured Data Analysis

Ethical Issues in Data Science and Big Data

Acquisition

Analysis

Action

Discussion

How might the policies of an organization in relation to document and content management raise ethical questions?

For example, if a government organization has been given the only copies of medical records for people who had undergone questionable medical procedures for the purposes of a redress scheme, would it be ethical for the organization to destroy those records rather than returning them to the survivors?

Discussion

Information security is a key function in the organization. Is it ethical to install keyloggers on company computers to have access to information about what staff are always doing in the organization? Why do you take that view?