DAMA Phoenix Event Presentation

Anatomy of a Data Request

Or “How many times do we have to do this?”

My Background

- Employed by TEKsystem (since 2020):
  - Data Architect Consultant
  - Focused on Data Architecture activities/practice and assisting/mentoring EDW team members
- Experience (IT, over 35 yrs; Data focus, over 20 yrs):
  - Disciplines:
    - Enterprise Data Strategy, Data Architecture, Data Design, Data Integration, Reference & Master Data, Data Warehousing, Business Intelligence, Data Governance, Metadata Management
  - Industries:
    - Insurance & Financial Services
    - Non-for-profit (membership & training)
    - Pharmaceutical
    - State Government
    - Manufacturing
    - Telecommunications
- Other Items:
  - Founding member (since 2009) and current President (2016) of DAMA Indiana chapter
  - Hold CDMP certification (Master level since 2010)
  - Contributing author to DM-BOK2 (Reference & Master Data)
  - Owner/principle of dataThink, LLC (2020)

Topic of Discussion

This is a case study of an actual data request put forth by a business functional area and submitted to IT for fulfillment.

This particular request was not overly complex, but the story of the fulfillment process exposed a number of "issues" that coalesced into a perfect storm.

<u>We will review the following:</u>

- Request fulfillment
- Issues encountered
- Lessons learned
- Suggested guidance
- Final Thoughts

(Note that the names have been changed to protect the guilty…☺)

Definition (from Oxford Dictionaries)

a·nat·o·my

NOUN

1. the branch of science concerned with the bodily structure of humans, animals, and other living organisms, especially as revealed by dissection and the separation of parts.
   - the bodily structure of an organism.
     "descriptions of the cat's anatomy and behavior"
     synonyms: structure, makeup, composition, constitution
   - informal
     a person's body.
     "every part of his anatomy hurt" ·
     synonyms: body, shape, figure, silhouette, proportions, statue, build
2. a study of the structure or internal workings of something.
   "Machiavelli's anatomy of the art of war"
   synonyms: analysis, examination, inspection, survey, study, scrutiny, perusal

Focus of our discussions

The Request...(names have been changed to protect the guilty)

- A regulatory area requested a subset of data from the organization's administrative systems
- The purpose was to corroborate existing processes within the organization
- The tenet of the request was to list accounts that were tied to a locality within a certain timeframe
- The type of data requested was normal, business-related data that typically exists within administrative systems*

	A	B	C	D
1	Account Number	Account Status	Effective Date	Account Locality
2	12345678	ACTIVE	01-01-2003	IN
3	22334455	ACTIVE	06-01-2002	OH
4	23412389	TERMINATED	02-30-2010	IN
5	23579135	ACTIVE	11-21-2013	IN
6	23-57913	ACTIVE	11-21-2013	IN
7	34267599	UNKNOWN	01-01-2003	IN
8	45454545	ACTIVE		IN
9	66117722	ACTIVE	08-01-2007	CA
10	72240911	ACTIVE	09-17-2008	OH
11	76540987	ACTIVE	01-01-2003	IN
12	88881234	ACTIVE	03-02-2001	IN
13	88881234	TERMINATED	03-20-2001	IN
14	98877665	ACTIVE	01-01-2003	IN

* Illustrative example

Initial Fulfillment of Request

Attempt #1

- Sent as regular 'Dat service desk ticket to be fulfilled by off-site IT support area

INCOMPLETE

Attempt #2

- Resent original service desk request was reviewed by on-site IT support area

STILL INCOMPLETE

Attempt #3

- Reviewed request with Business iar with these type of requests

INCORRECT

Attempt #4

- Request was further clarified by Busin some profiling by IT support folks

STILL INCORRECT

Escalated Fulfillment of Request

Attempt #5

- Business area be[illegible]ation and asks for assistance to understand why request [illegible]being fulfilled correctly

ISSUES STILL EXIST

Attempt #6

- Decision by IT area (sp[illegible]lly the CIO) to establish SWAT to review request and pull in mo[illegible]ources to help

NOT REALLY

Attempt #7

- Senior leadership from Busi[illegible]T get involved and importance of request is…ummm…[illegible]ied

? GETTING CLOSER

Completion of Request

Attempt #8

- Additional profi[illegible] source systems helped identify missing content from previous [illegible]quests

STILL ISSUES

Attempt #9

- Semantic issues rec[illegible]ong request, source systems, and compiled results set

? GETTING CLOSER

Attempt 10#

- Request reviewed on-site [illegible]gulatory area representatives

✓ WELL, MAYBE

Completion of Request

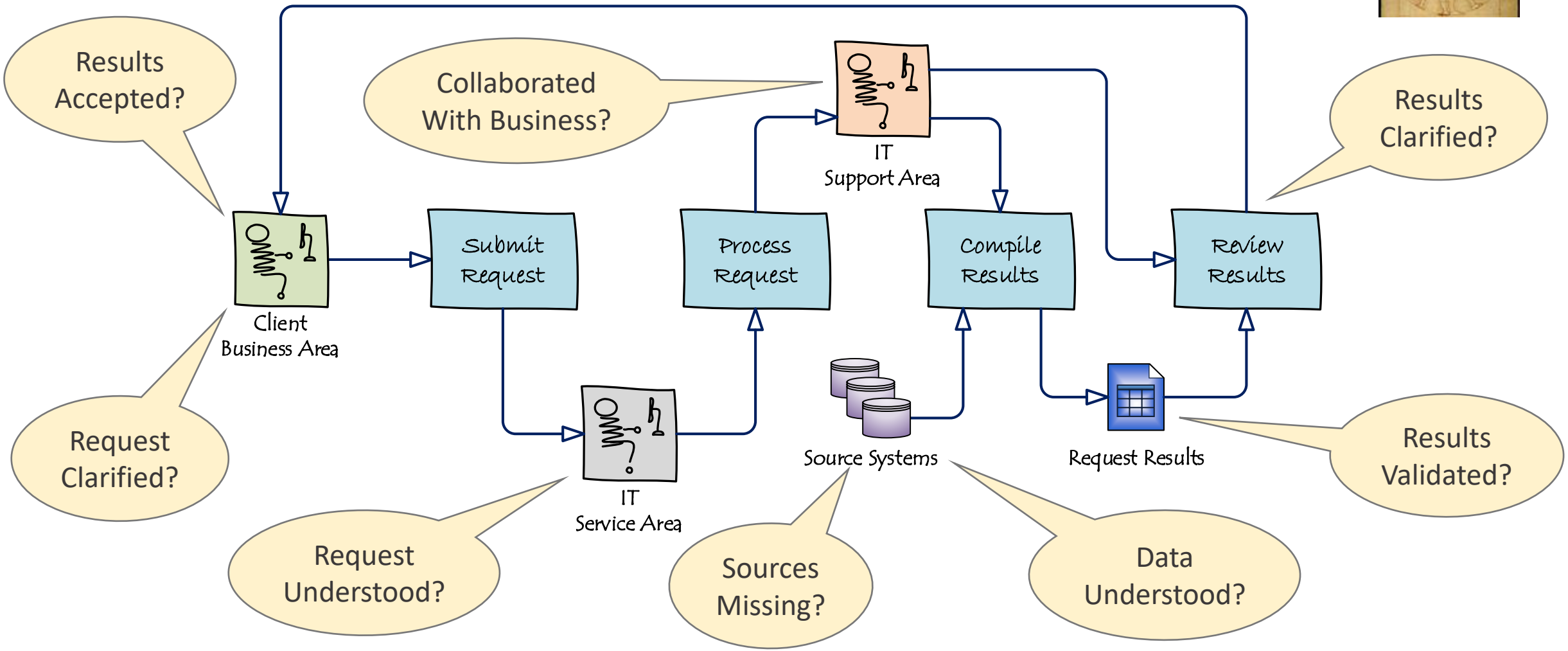
Still not sure if request was ever completed 100% per the requirement

Here are some stats (approximations) surrounding the request:

- Number of Attempts: 10
- Elapsed Time: 14 weeks
- Number of Participants: 6 average; 20 peak
  - (includes Business, IT, and external resources)
- Effort Cost: $250k*
- Expected Cost: $1k

*Cost of travel, many lunches, and aspirin not included…☺

What Went Wrong – Process



What Went Wrong – Data

	A	B	C	D
1	Account Number	Account Status	Effective Date	Account Locality
2	12345678	ACTIVE	01-01-2003	IN
3	22334455	ACTIVE	06-01-2002	OH
4	23412389	TERMINATED	02-30-2010	IN
5	23579135	ACTIVE	11-21-2013	IN
6	23-57913	ACTIVE	11-21-2013	IN
7	34267599	UNKNOWN	01-01-2003	IN
8	45454545	ACTIVE		IN
9	66117722	ACTIVE	08-01-2007	CA
10	72240911	ACTIVE	09-17-2008	OH
11	76540987	ACTIVE	01-01-2003	IN
12	88881234	ACTIVE	03-02-2001	IN
13	88881234	TERMINATED	03-20-2001	IN
14	98877665	ACTIVE	01-01-2003	IN

Effective Date, of what?

Incorrect Account?

Locality now, or when Account set up?

Bad Date?

Invalid Status?

Locality refers to California, or Canada?

Duplicate Accounts, or same Account?

Missing Data?

What Was Lessons Learned*

Various failures of both process and data…

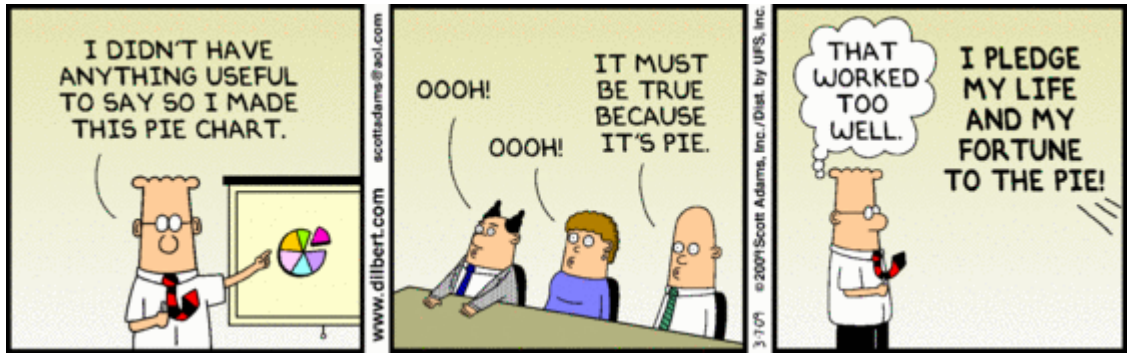
- Process related:
  - No clarification or classification of request
  - Just shuffled to IT support without any expectations of outcome desired
  - No collaboration or communication with business partners
  - Importance of request not understood

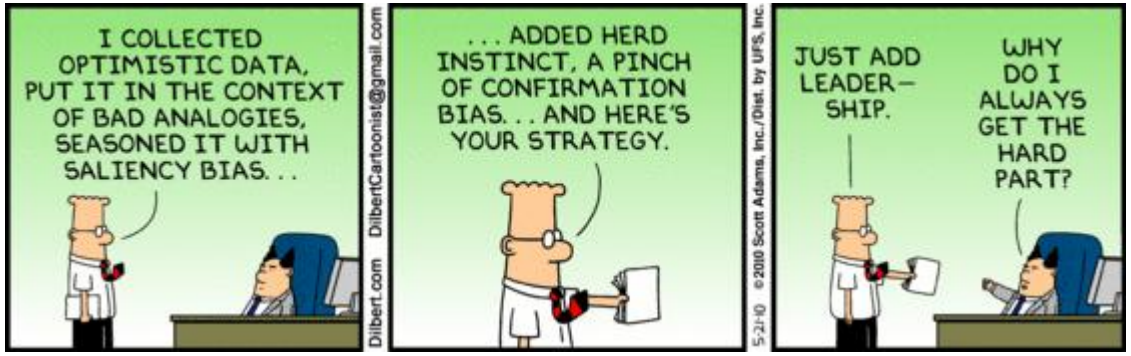
- Data related:
  - Data items not clearly defined in request
  - Minimal understanding of source systems
  - Semantic differences of data items
  - No approach to validate results

* initial review, with expected changes to be put forth later

Need a Break…here's some Dilbert…

This is kinda dated…but I like pie!





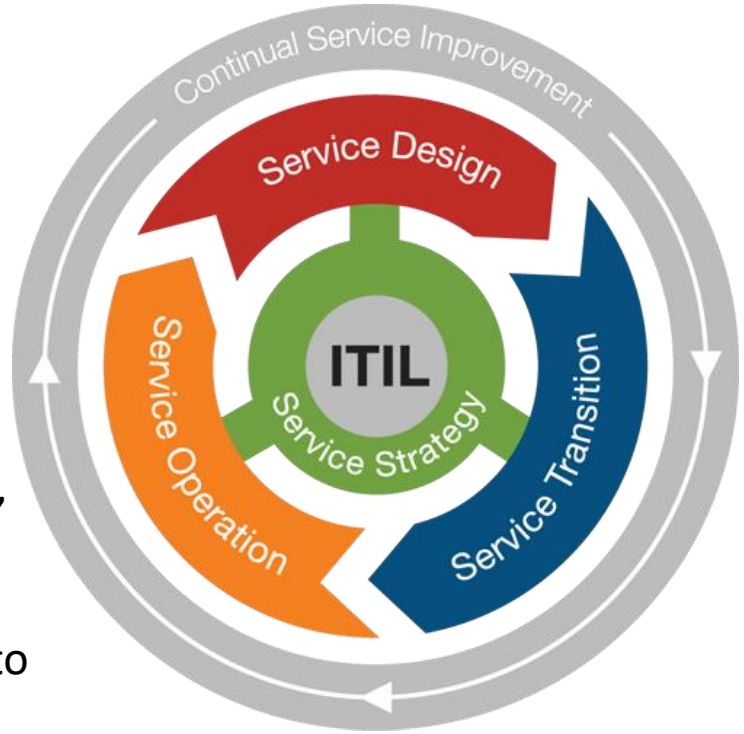
←Replace Dilbert's "I" with "AI"…hmmm

…but what are some things that could help?

What Can Help – Process (ITIL*)

ITIL (Information Technology Infrastructure Library) is a globally recognized framework for IT service management (ITSM) that provides best practices to align IT services with business needs. It helps organizations deliver high-quality IT services efficiently, improve customer satisfaction, and manage processes like incident resolution, change management, and service delivery. Core Components include:

- **Service Value System (SVS)** – Ensures all components and activities of the organization work together to create value.
- **Four Dimensions of Service Management** – Focuses on:
  - Organizations and People: Structure, roles, and culture.
  - Information and Technology: Tools, data, and systems.
  - Partners and Suppliers: External relationships and outsourcing.
  - Value Streams and Processes: Workflows to deliver value.
- **Guiding Principles** – Like “focus on value” and “think and work holistically” to guide decision-making.
- **Service Management Practices** – 34 practices grouped across General (strategy, risk), Service (incident, change control), and Technical (deployment, platform) categories.



Refactoring the existing **request process** to follow tenets of ITIL will be beneficial to successful fulfillment of future data requests.

* originally developed by the UK government in the 1980s, now maintained by **Axelos** (ITIL 4, the latest version)

What Can Help – Process (DCLC*)

The **Data-Centric Development Life Cycle (DCLC)** is an iterative methodology that emphasizes addressing data issues and opportunities early in the development process, rather than during testing or after production implementation. This approach contrasts with traditional Systems Development Life Cycle (SDLC) models, which often prioritize process and application development over data considerations.

- **Early Data Focus**: DCLC brings data problems and opportunities to the forefront before development begins, ensuring that data quality and governance are integral from the start.
- **Iterative Process**: The methodology supports continuous refinement, allowing for adjustments as data requirements and business needs evolve.
- **Integration with Existing Methodologies**: While DCLC can stand alone, it can also complement traditional SDLC or Agile approaches by embedding data-centric practices within these frameworks.

By adopting a data-centric approach, organizations can ensure that data considerations are not an afterthought but a foundational element of system development.

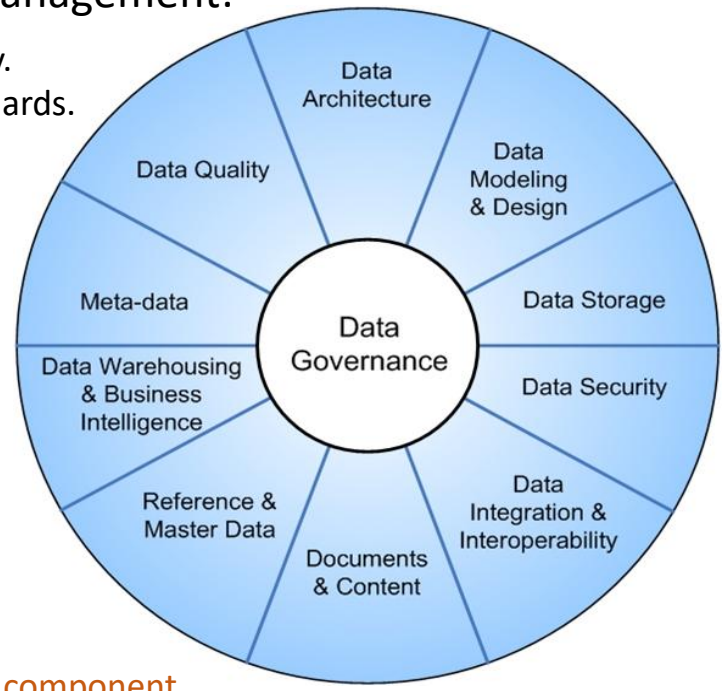
Following this process leads to more **robust, reliable, and compliant data systems** that better serve future data requests.

* Malcolm Chisholm – First San Francisco Partners

What Can Help – Data (DMBOK*)

The **DAMA DMBOK** (Data Management Body of Knowledge) outlines 11 core **data management knowledge areas**. These components form the framework for effective data governance and management:

1. **Data Governance** – The overarching function that guides data management and ensures accountability.
2. **Data Architecture** – The design and maintenance of the data environment, including models and standards.
3. **Data Modeling and Design** – Techniques for analyzing and designing data structures.
4. **Data Storage and Operations** – Management of data storage, access, and lifecycle.
5. **Data Security** – Ensuring data privacy, protection, and risk mitigation.
6. **Data Integration and Interoperability** – Combining data from different sources and ensuring it works together effectively.
7. **Document and Content Management** – Managing unstructured data like documents and multimedia.
8. **Reference and Master Data Management (MDM)** – Managing consistent and accurate reference and master data across systems.
9. **Data Warehousing and Business Intelligence (DW/BI)** – Supporting decision-making with structured data and analytics.
10. **Metadata Management** – Managing data about data to improve understanding and usage.
11. **Data Quality Management** – Ensuring data is accurate, complete, and fit for use.
12. Artificial Intelligence – There is a **Big Data** section in the DMBOK2, perhaps this expands/becomes "AI" component.



Understanding the DMBOK core components will make a qualitative difference in **gathering data results** for future data requests.

* DAMA International - dama.org (DMBOK2 current version, DMBOK3 currently being worked on by DAMA-I)

What Can Help – Data (Maturity)*

Data maturity refers to an organization's ability to effectively collect, manage, analyze, and utilize data to drive decision-making and achieve business goals. It reflects the sophistication and integration of data processes, technologies, and culture within the organization. Data maturity is often assessed across 5 stages:

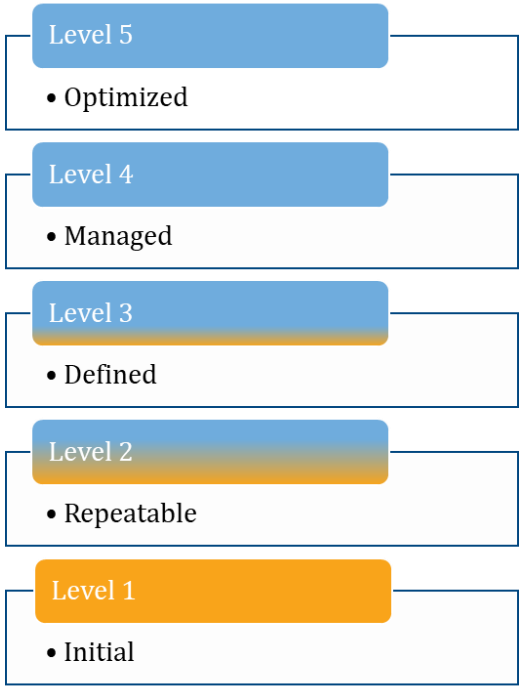
1.Initial (Ad Hoc / Chaos)
- Data is everywhere but nowhere, no formal data management practices, high risk with using data
- Mindset: "We have data, but it's messy and nobody trusts it."

2. Repeatable (Developing / Awareness)
- Early efforts to manage data appear, some departments own and manage their data, pain points are recognized
- Mindset: "We know we have a problem... and we're starting to fix it."

3. Defined (Standardized)
- Formal processes for managing data emerge, roles (i.e. Data Steward and Data Owner) are assigned, basic data governance framework in place, data definitions are standardized across business units.
- Mindset: "Data is becoming an asset, not a burden."

4. Managed (Measured and Controlled)
- Data quality is monitored, metadata management and master data management (MDM) practices exist, data is integrated across systems, compliance with regulations like GDPR or HIPAA becomes systematic.
- Mindset: "We trust our data to make decisions and run our operations."

5. Optimized (Innovative / Predictive)
- Data is a strategic advantage, advanced analytics, AI, and machine learning drive predictions, continuous improvement of data processes, employees at all levels think about and use data responsibly and creatively.
- Mindset: "Data drives everything we do — from insight to innovation."



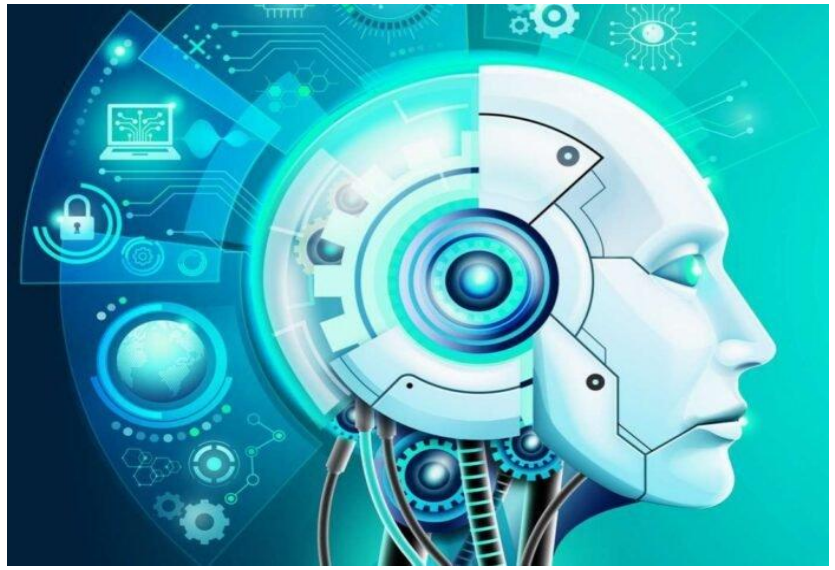
Higher data maturity correlates with better **decision-making, operational efficiency, and customer insights** for future data requests.

* there are various data maturity models available (e.g. DAMA-DMBOK, Gartner, CMMI, IBM, DCAM, Microsoft)

What Can Help – Artificial Intelligence*

AI enhances overall Data Management by automating and optimizing key processes, improving efficiency, accuracy, and scalability. Here's how AI contributes across various aspects:

- **Data Collection**: AI automates data extraction from diverse sources (e.g., documents, web, IoT devices) using techniques like natural language processing (NLP) or computer vision, reducing manual effort and errors.
- **Data Storage**: AI optimizes storage by predicting data growth, recommending cost-effective solutions (e.g., cloud vs. on-premises), and automating data archiving or tiering based on usage patterns.
- **Data Organization**: AI-powered tools classify and tag data (e.g., metadata generation) using machine learning, enabling better indexing and retrieval. It can also detect relationships in unstructured data for improved structuring.
- **Data Quality Control**: AI identifies inconsistencies, duplicates, or missing values in datasets, suggesting or auto-correcting errors. Anomaly detection algorithms flag outliers, ensuring data reliability.
- **Data Security**: AI enhances security by detecting threats in real-time (e.g., unusual access patterns), automating encryption, and predicting vulnerabilities. It also supports compliance with regulations like GDPR through automated audits.
- **Data Governance**: AI enforces governance policies by monitoring data usage, ensuring compliance, and generating reports. It can also recommend policies based on regulatory changes or organizational needs.
- **Data Analysis/Utilization**: AI enables advanced analytics (e.g., predictive modeling, sentiment analysis) and powers applications like recommendation systems or chatbots, extracting actionable insights from raw data.



Benefits: AI reduces manual workloads, scales operations, and uncovers insights that humans might miss. For example, tools like automated ETL (Extract, Transform, Load) pipelines or AI-driven data catalogs streamline workflows.

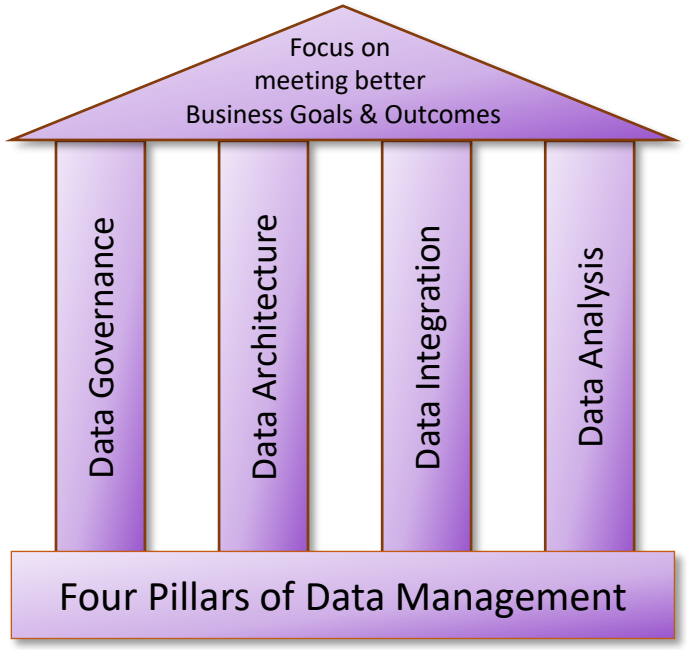
Challenges: AI implementation requires investment, skilled personnel, and careful oversight to avoid biases or errors in automated processes.

By integrating AI, organizations can manage data more proactively, securely, and efficiently, driving better decision-making and innovation…**for future data requests**.

* the information on this slide came directly from Grok 3

Focus on Data Management

To "focus" on Data Management means prioritizing the strategic and effective handling of data throughout its entire lifecycle to meet organizational goals and outcomes. This involves dedicating attention, resources, and expertise to ensure that data is accurate, secure, accessible, and valuable. If the DMBOK components were grouped across “Four Pillars of Data Management”, it would look like the following:



1. Data Governance
- **Data Governance** (The core function ensuring proper oversight, policies, and stewardship across all areas.)
- **Data Quality Management** (Ensures data is accurate, complete, and reliable, which is central to governance.)
- **Data Security** (Protects data integrity, privacy, and regulatory compliance aspects of data.)

2. Data Architecture
- **Data Architecture** (Designs and maintains the structure of the data environment.)
- **Data Modeling and Design** (Builds data models, schemas, and logical structures based on business needs.)
- **Reference and Master Data Management (MDM)** (Ensures consistency of key entities across systems.)

3. Data Integration
- **Data Integration and Interoperability** (Combines and synchronizes data from multiple sources.)
- **Data Storage and Operations** (Manages the infrastructure and lifecycle of data storage and access.)
- **Document and Content Management** (Manages unstructured data like documents, media, and emails.)

4. Data Analysis
- **Data Warehousing and Business Intelligence (DW/BI)** (Supports analytics and reporting to aid business decisions.)
- **Metadata Management** (Provides context and lineage for data utilization, enabling transparency and compliance.)
- **Artificial Intelligence (AI)** (Applies advanced algorithms to automate insights and predictions.

These pillars provide a strong foundation for effective data management and help organizations leverage data as a strategic resource.

Focus on Business Outcomes

To "focus" on Business Outcomes means directing efforts, resources, and decision-making toward achieving tangible, measurable results that align with an organization's strategic goals. Instead of concentrating **solely on processes or tools**, the emphasis is placed on the **end results** that contribute to growth, efficiency, or innovation. This involves:

Clarifying Objectives: Identifying the specific results to achieve (e.g. increased revenue, customer satisfaction, or market share.)

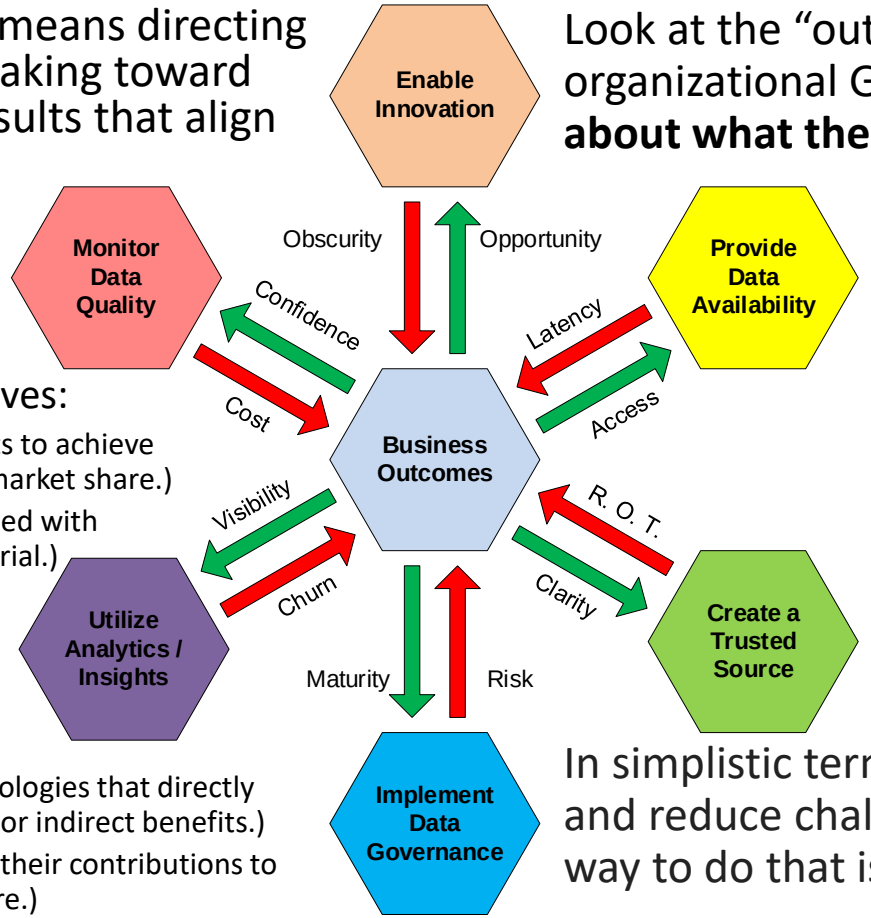
Aligning Strategies: Ensuring all initiatives are aligned with objectives (operational, technological, and managerial.)

Measuring Impact: Establishing key performance indicators (KPIs) to track progress.

Adapting to Change: Remain flexible and adjust strategies (e.g. evolving market demands or organizational priorities.)

Prioritizing Value: Investing in initiatives and technologies that directly contribute to business outcomes (ignore uncertain or indirect benefits.)

Driving Accountability: Encouraging teams to own their contributions to business outcomes (create a results-oriented culture.)



Look at the “outer” hexagons as possible organizational Goals (**broad, general statements about what the organization aims to achieve**) tied to Business Outcomes at the center (**specific, measurable results indicating progress, impact on those goals**.)

The **green** arrows denote the outcome benefits (*what you want to increase, gain*) to the goals whereas the **red** arrows denote the outcome challenges (*what you want to reduce, remove*) of the goals.

In simplistic terms, you want increase benefits and reduce challenges of stated goals. The best way to do that is focus on Business Outcomes.

Outcome of the Data Request*

- Requirements Definition
  - Request process triage
  - Coverage of attributes requested
  - Scope of data sources requested
  - Request negotiation & response
- Data Collection
  - Data extraction patterns
  - Data integration architecture
  - IT validation process
- Business Validation
  - Testing scope
  - Sampling techniques
  - Package & delivery process
- Gap Recommendations
  - Creation of Data Operations team
  - Technology review

*based on post-mortem review by Business and IT partners

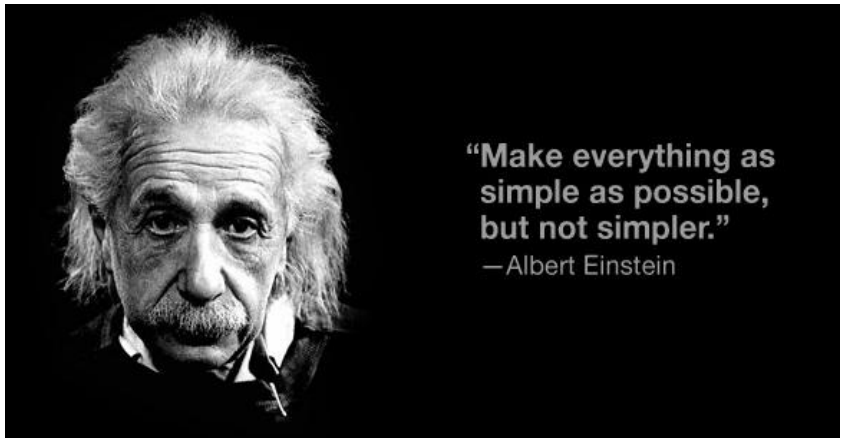
Final Thoughts

First off, this discussion on this single Data Request may appear extreme or more of an anomaly…

However, this occurs more often than we think…is just a matter of degree.

This incident illustrates the importance of having **process** and **data** framework(s) available, established, understood, and utilized.

Following even the smallest tenets of ITIL, DCLC, and DMBOK would have resulted in a better outcome…IMHO



It may seem overwhelming, but starting small and implementing certain aspects of Data Management will succeed and can be built upon. To quote another famous character, "just keep swimming…just keep swimming"…☺

And the hope is to learn from the past mistakes and not repeat them!

Questions

Thank you for your time and interest

Contact Information:

Gene Boomer

dataThinkLLC

gene@datathinkllc.com