

# Data Management

## Best Practices

### Practicing Data Management Better



[peter.aiken@anythingawesome.com](mailto:peter.aiken@anythingawesome.com) +1.804.382.5957



1 + 1 = 11



© Copyright 2023 Peter Aiken, PhD Slide # 1

### Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS ([vcu.edu](http://vcu.edu))
- Institute for Defense Analyses ([ida.org](http://ida.org))
- DAMA International ([dama.org](http://dama.org))
- MIT CDO Society ([iscdo.org](http://iscdo.org))
- Anything Awesome ([anythingawesome.com](http://anythingawesome.com))
- Experienced w/ 500+ data management practices worldwide
- Multi-year immersions
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - HUD ...

**\$1,500,000,000.00 USD**



- 12 books and dozens of articles



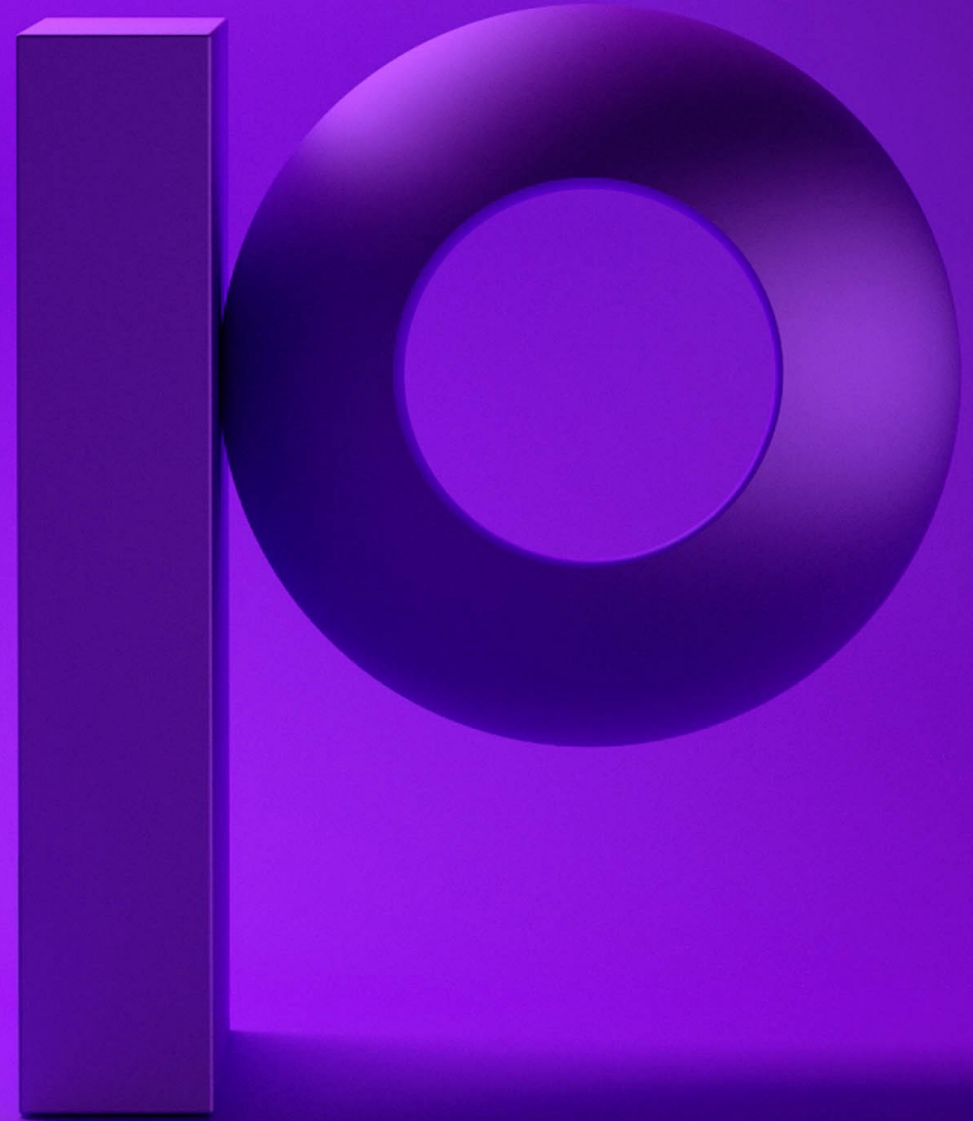
<https://anythingawesome.com>



precisely

# Data Management: Data Integrity as a Journey

Hosted by Dataversity



# Macro forces are impacting your business



---

Long-term  
pandemic impacts



---

Changing  
regulatory and privacy  
environment



---

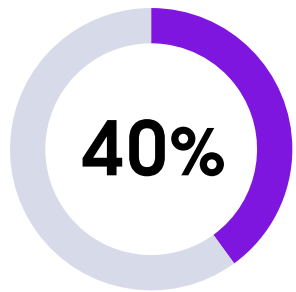
Volatile  
market conditions



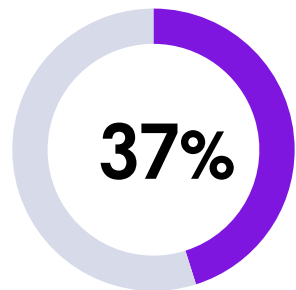
# Economic downturn...

QUESTION:

How has the economic downturn impacted your organization's data strategy?



Decrease in staffing & resource allocation



Decrease in budget

# ... impacts data strategy

QUESTION:

Which of the following trends are influencing your organization's overall data strategy?

57% Cloud adoption

41% AI/ML

49% Advanced data analytics

30% DataOps

43% Workflow automation

21% Data democratization

43% Digital transformation




Respondents are turning to cloud, advanced analytics, and automation to drive down costs and resource requirements



77%

report data-driven  
decision-making as the  
top priority for data  
programs



*2023 Data Integrity Trends & Insights*

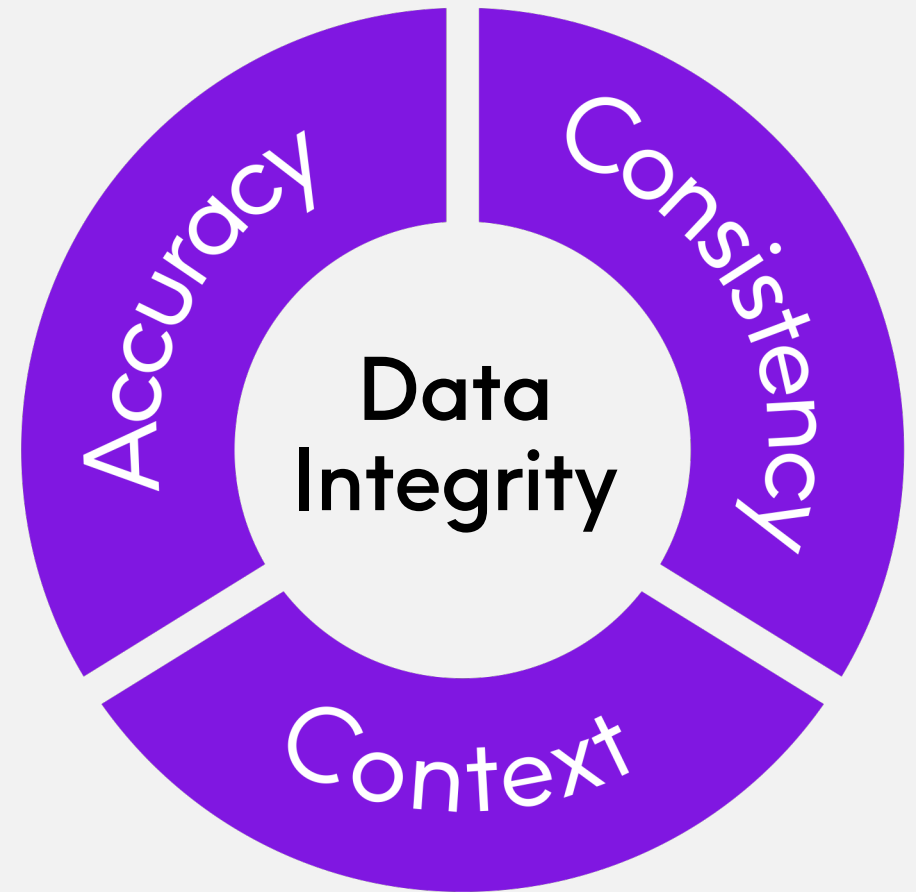
# Business-first approach

- Identify and prioritize the data that matters
- Link data assets to business goals
- Build stakeholder engagement across three levels (strategic, operational & tactical)
- Embed strategy in everyday business tasks



# For trusted data, you need data integrity

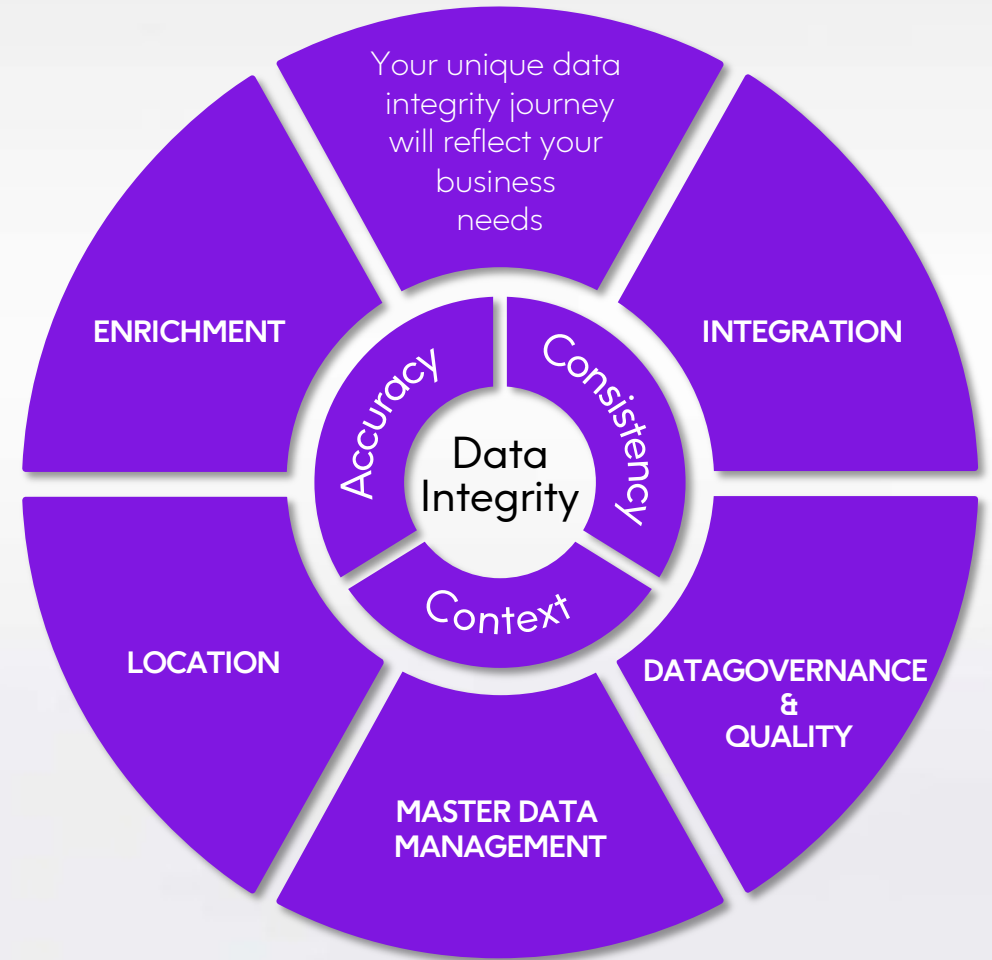
Data integrity is data with maximum accuracy, consistency, and context for confident business decision-making



# Data integrity is a journey

- Every journey to data integrity is unique and driven by business initiatives
- Market trends are accelerating the need for data integrity
- Precisely can help you at every step of your data integrity journey

**Strategic Services can help jump-start your initiative or ensure an existing one accomplishes the desired outcomes!**

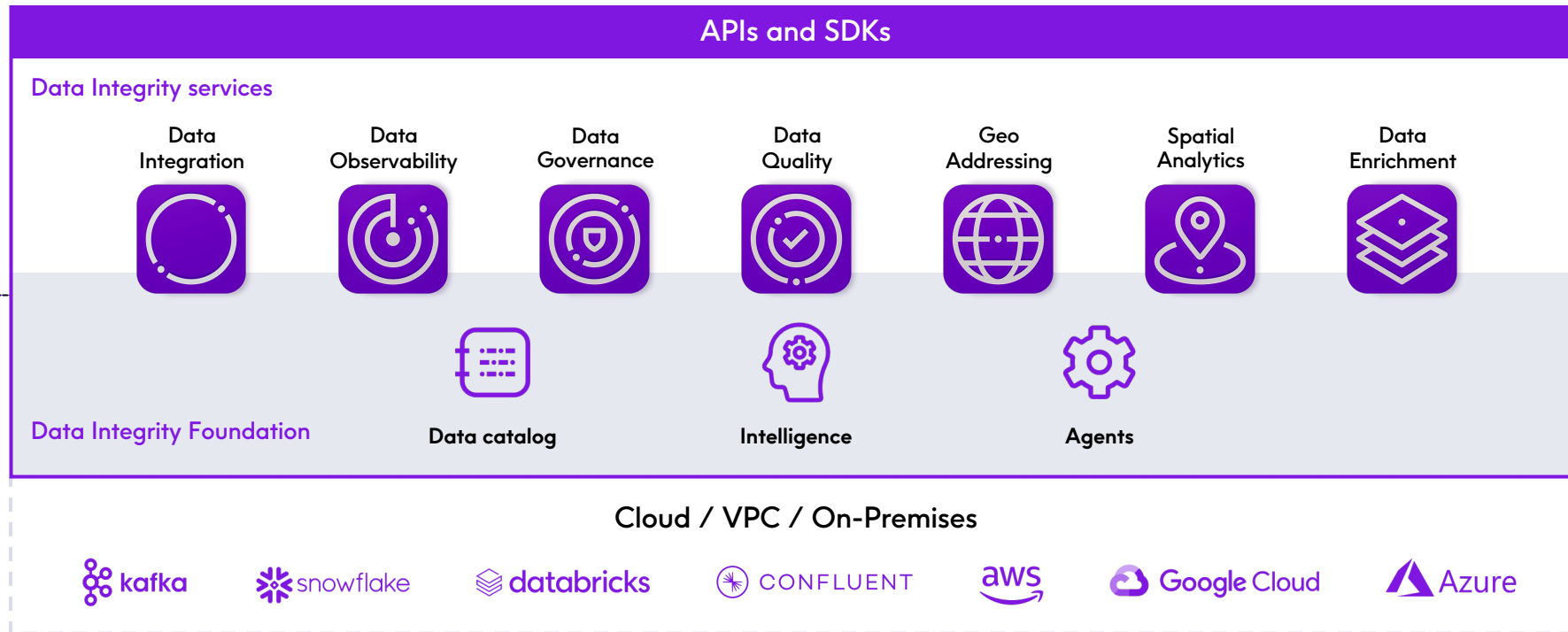




# Data Integrity Suite

## Enterprise Data Sources

- Business Intelligence
- CRM
- Workforce mgmt.
- Data warehouse
- ERP
- Billing



## Enterprise Business Systems

- Enterprise apps
- Analytics tools
- Precisely industry apps
- BI dashboards
- AI/ML



## The leader in data integrity

Our software, data enrichment products and strategic services deliver accuracy, consistency, and context in your data, powering confident decisions.

**12,000**  
customers

**99**  
of the Fortune 100

**100**  
countries

**2,500**  
employees

### Brands you trust, trust us



### Data leaders partner with us



# Let's continue the conversation...



[www.precisely.com](http://www.precisely.com)

Demos  
White Papers  
Case Studies



**Contact us**

Set up a 30-minute no-obligation consultation with our team  
+1-877-700-0970

“Get in touch” on [www.precisely.com](http://www.precisely.com)

**precisely**





# Data Security Platform

## Set Your Data Free

December 2023

//

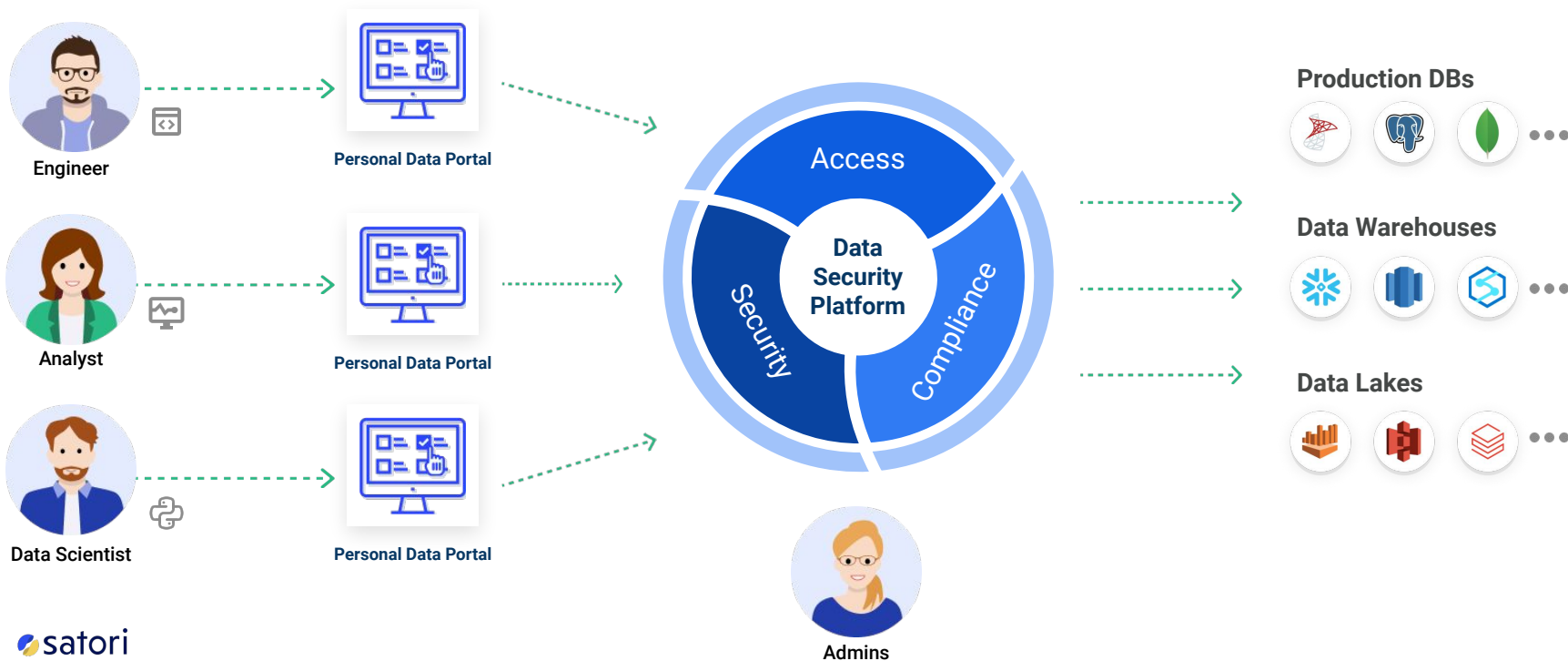
You need to make secure data the fastest way of getting things done and that's what we do with Satori.



**Dr. Diederik Van Liere**

VP Data Science and Engineering, *Wealthsimple*

Satori is a **Data Security Platform** that provides **immediate** and **secure** access to data for **everyone** in your company.



# What it takes to get data to people.

## Access

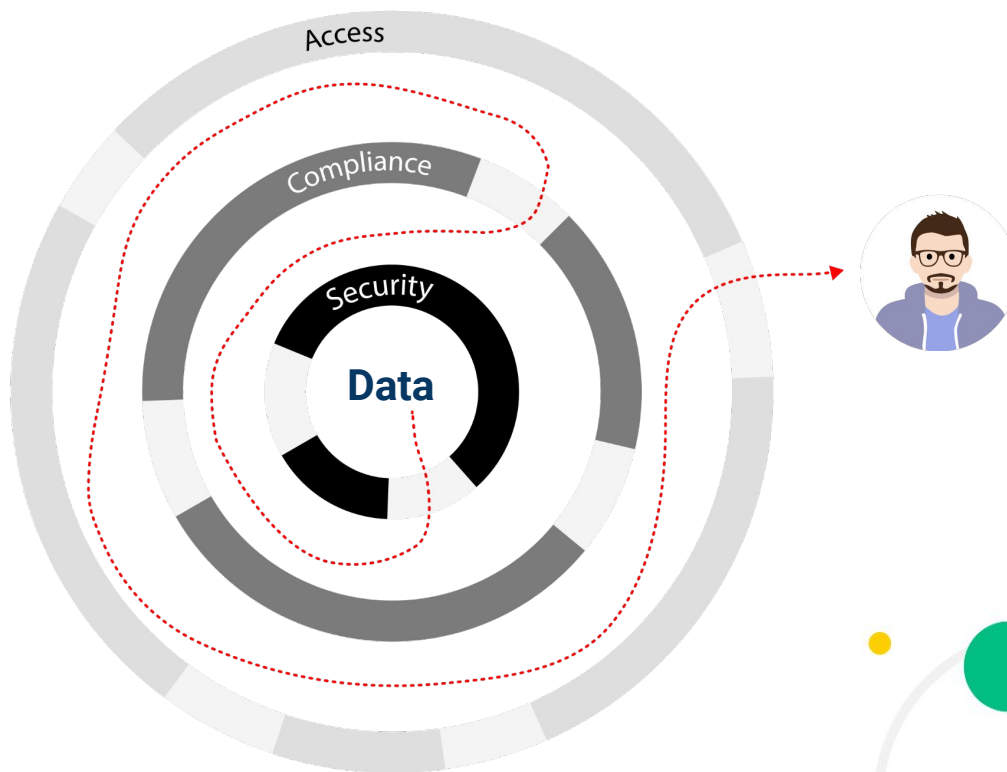
- RBAC
- Access Requests
- Access Reviews
- Permissions Management

## Compliance

- GDPR
- SOC2
- SOX
- HIPAA
- DPAs
- PII Protection
- Audits
- Classification
- Privacy Reviews

## Security

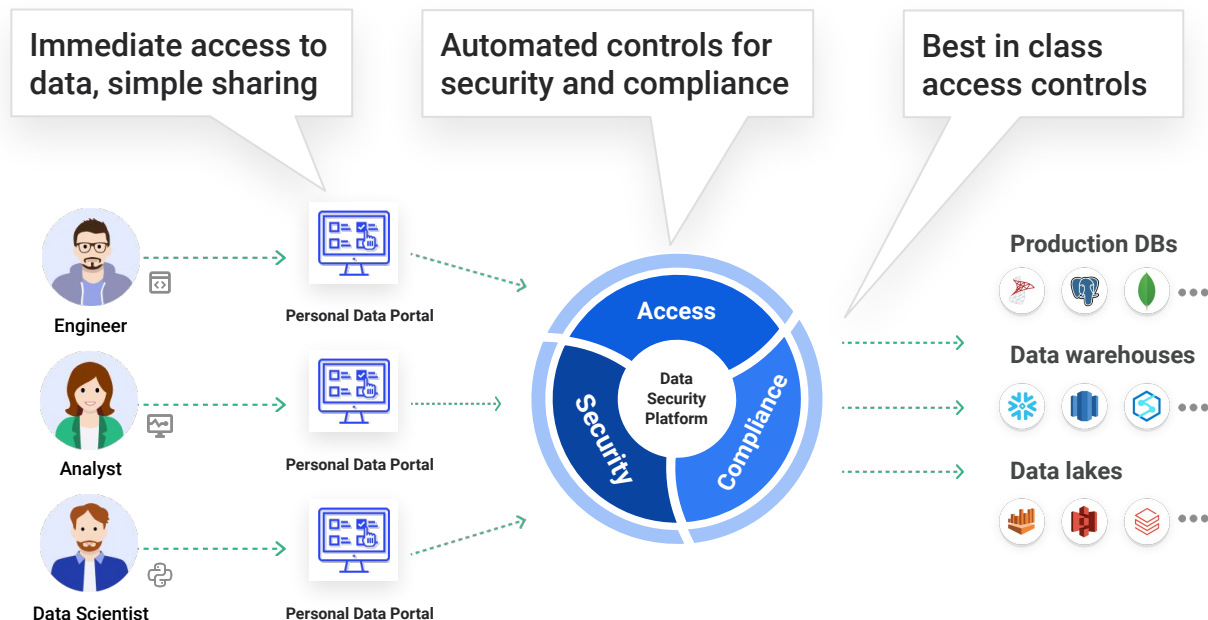
- Monitoring
- Classification
- Policies
- Detection
- Forensics
- Masking
- Security Reviews



# Before Satori

- It takes a long time to get access to data
- Data is siloed, and sharing is difficult
- Large investment in manual security and compliance
- Less than desirable access controls

# With Satori



# Turning Data to Value. Simple?

Hours

## Discover

Where the data is.  
Who has access to it.  
Sensitive data.  
Data usage.

### Typical integrations

IdP (Active Directory, Okta)  
Databases, warehouses, lakes

Days

## Define

Datasets.  
Access workflows.  
Policies.

### Typical integrations

BI tools (Tableau, PowerBI)  
Query tools (Redash, DBeaver)  
Workflow (Jira, ServiceNow)

Up & Running

## Deliver

Personal data portal.  
Universal security and compliance.  
Automated access and policy creation.

### Typical integrations

SIEM (Splunk, Datadog)





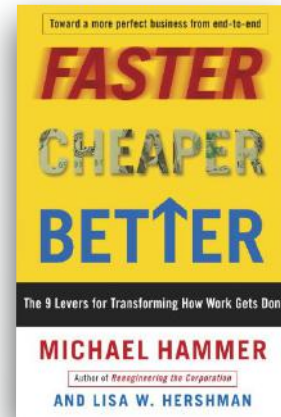
# Thank you!

[contact@satoricyber.com](mailto:contact@satoricyber.com)

[www.satoricyber.com](http://www.satoricyber.com)

# Doing Data Better means that you

- Understand that the vastness and quality of data plays an increasing role in everyone's life
- Are motivated to increase your individual data skills because you now know that poor data skills:
  - Cost you more
  - Steal increasing amounts of your time
  - Deliver less
  - Presents greater risk
- Recognize the critical importance of data management in modern life and its positive and negative applications
- Develop defensive skills to differentiate between good and bad data (understanding that most data is of unknown quality)
- Can assign values to some of your personal data and its use
- Are able to take advantage of decreases in the general workload load needed to effectively manage data in your professional and personal life



## Program overview

- Motivation
  - Frustration—we are unsatisfied with current state
  - Are we making progress? (No)
- How did we get here? (Building on proven research)
  - DoD → SEI → MITRE → CMMI
  - Industry push for best practices
- Ingredients
  - What is the Data Maturity Model? (DMM)
  - Body of Knowledge (DM BOK)
- Understanding and applying them together
  - Weakest link in the chain architecture
  - Just a bit on strategy
  - Three legged stool
  - How does one get to Carnegie Hall?
- Where to next?
- Q & A?

Practicing  
Data Management  
Better



1 + 1 = 11



## Measures of Unproductivity

### Knowledge Worker Stress

- 33% of time spent reworking/ recreating knowledge that already exists!
- 10% of time spent creating new knowledge and content
- 53% would rather to household chores
- 52% would rather pay bills than use content management/repositories
- 74% report feeling **overwhelmed** or **unhappy** when working with data
- 33% of overwhelmed employees spend at least one hour a week **procrastinating** over data-related tasks



## Measurements

### Everyone

- **14%** have a good understanding of how to use business data
- **21%** aged 16-24 classified themselves as data literate

**Conclusion:** future employees are underprepared for data-driven workplaces

### Business decision makers

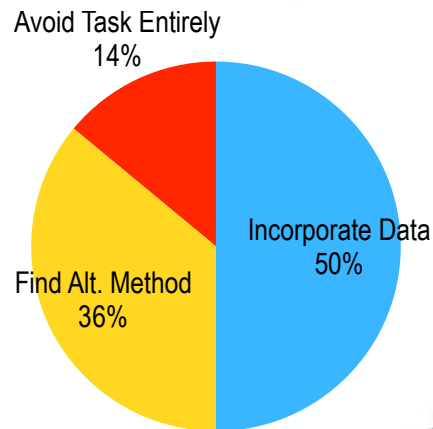
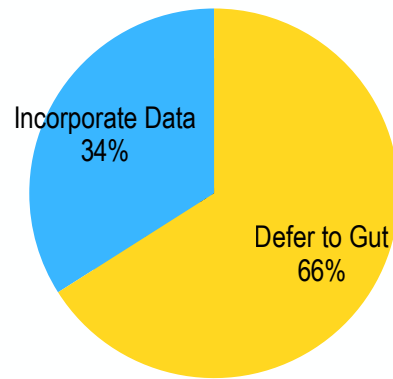
- **24%** of business decision makers feel fully confident in their ability to read, work with, analyze & argue with that data
- **33%** are able to create measurable value from data
- **27%** say my analytics projects produce actionable insights
- **78%** willing to invest time/energy improving data skillsets



## When asked to incorporate data

**Too many organizations have simply put data in the hands of employees and expected them to make a success of it**

- Data appreciation isn't translating into employee adoption
  - 48% frequently make gut decisions
  - 66% for C-suite executives
- Lack of data skills is limiting workplace productivity
  - 36% said they would find an alternative method to complete the task **without using data**
  - 14 percent **avoid the task entirely**



https://anyorganization.com

http://TheDataLiteracyProject.org



© Copyright 2020 by Peter Allen Slide # 7

## Why weren't my data problems solved when we

- moved to the cloud?
- built the data warehouse/lakehouse?
- invested in technology?
- hired a CDO?
- purchased Salesforce?



https://anyorganization.com

© Copyright 2020 by Peter Allen Slide # 8

Data in the cloud should have three attributes that data outside the cloud/warehouse should not have. It should be:

# Cleaner

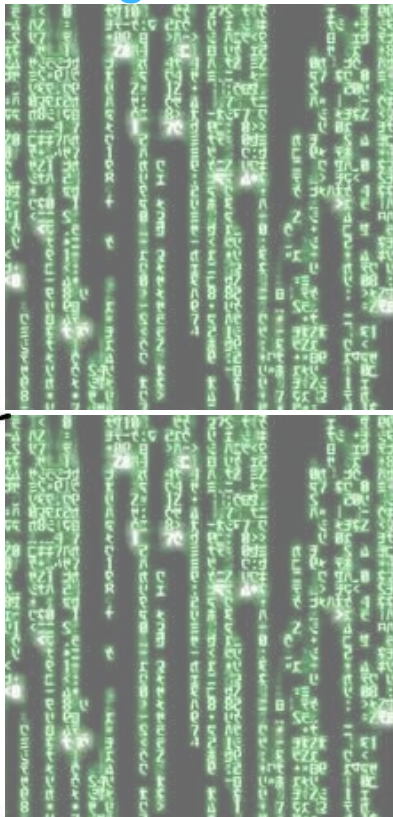


# Smaller

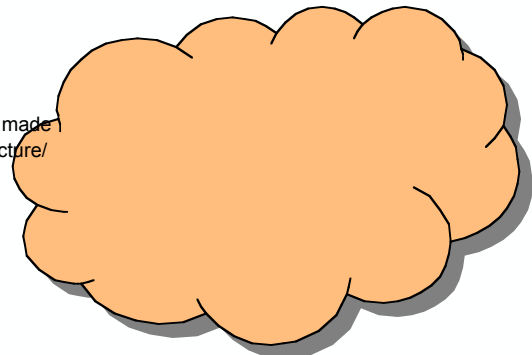
# Sharable-er



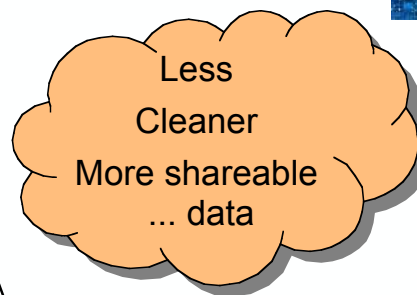
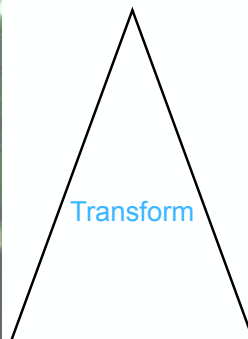
## Making ~~W~~Clouds ~~s~~uccessful



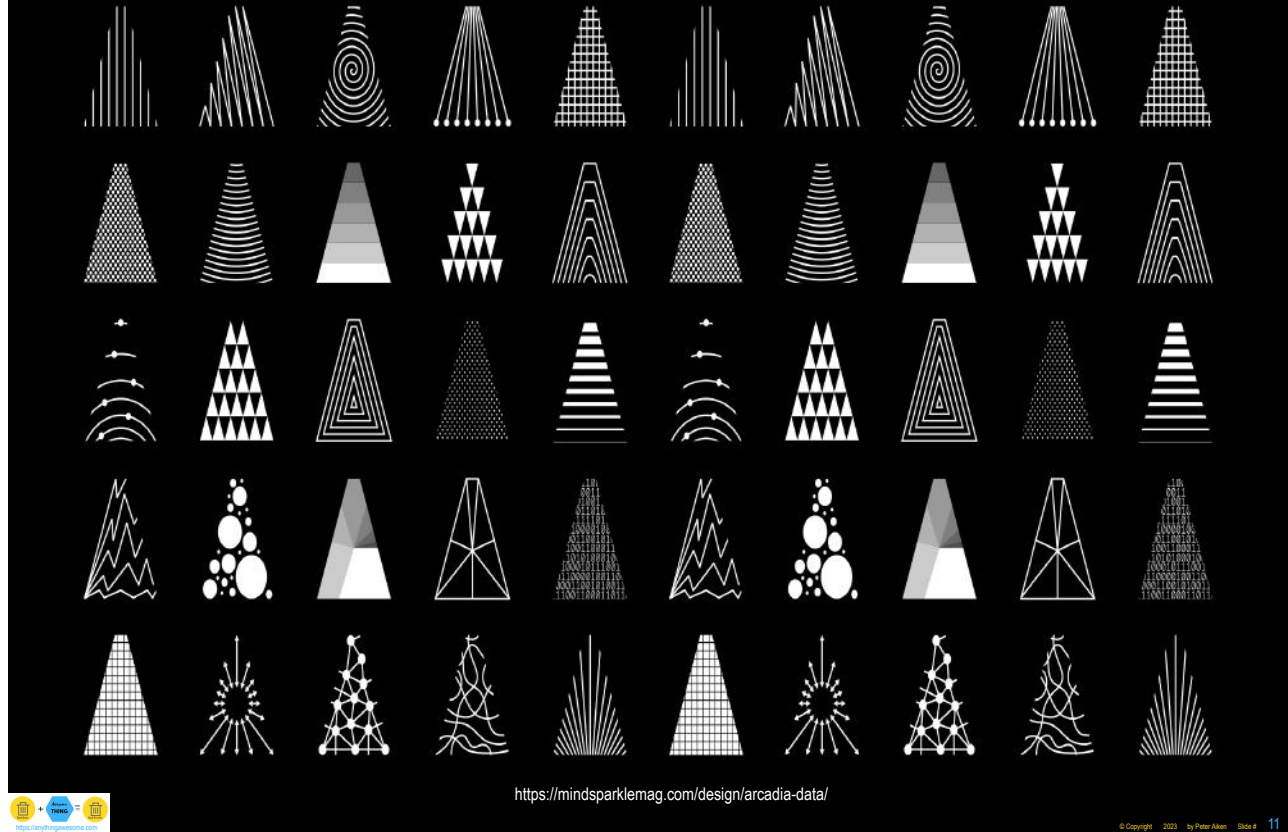
- Problems with forklifting
1. no basis for decisions made
  2. no inclusion of architecture/engineering concepts
  3. no idea that these concepts are missing from the process
  4. 80% of organizational data is ROT



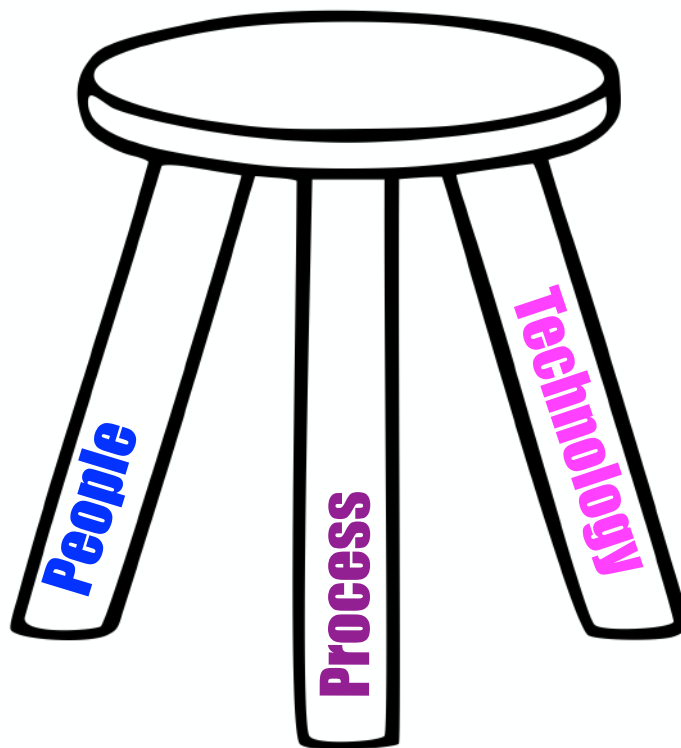
Data Branding



# Data Branding



# Success Requires a 3-Legged Stool



# Current approaches are not and have not been working

## What We Learned From Top Execs About Their Big Data And AI Initiatives

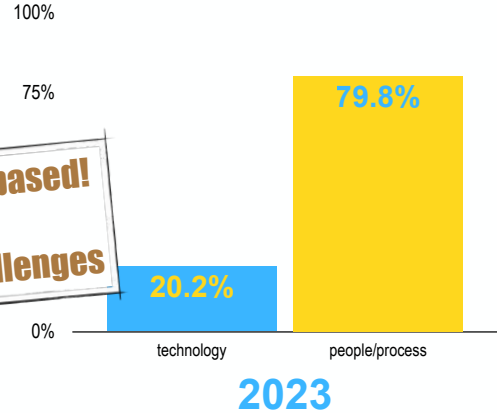


Randy Bean Contributor  
CIO Network Contributor Group @  
Enterprise & Cloud

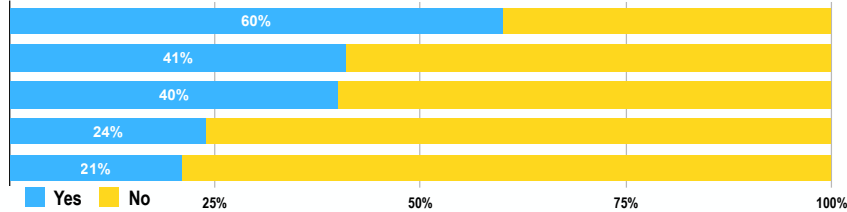
### Culture still eats strategy for breakfast

This aphorism is attributed to legendary management consultant Peter Drucker, and it certainly appears to hold true for data transformation efforts. Leading companies continue to identify culture – people, process, organization, change management – as the biggest impediment to becoming data-driven organizations – 92.2%. Few companies – only 30.0% – have even developed a well-articulated data strategy that culture could eat for breakfast. There is plenty of work to be done.

**80% of data challenges are people/process based!  
&  
DG is the only resource to address these challenges**



- Driving Innovation with Data
- Competing on data and analytics
- Managing data as a business asset
- Created a data-driven organization
- Forged a data culture



Source: Big Data and AI Executive Survey by Randy Bean and Thomas Davenport: [www.newvantage.com](http://www.newvantage.com)

© Copyright 2023 by Peter Allen Slide 13

Sample from: <https://artist.com/kathy-linden/on-outside-looking-in/?artic=4385>



## External Comprehension

Everything Else Data

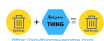
Data Management

Data Governance Program

Data (blah blah blah)

Most do not appreciate the difference between Data Governance and the other data stuff that needs to be done

# Data Program

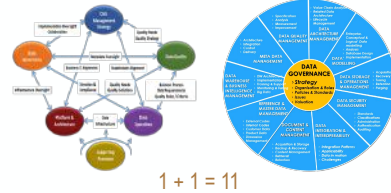


© Copyright 2023 by Peter Allen Slide 14

# Program overview

- Motivation
  - Frustration—we are unsatisfied with current state
  - Are we making progress? (No)
- How did we get here? (Building on proven research)
  - DoD → SEI → MITRE → CMMI
  - Industry push for best practices
- Ingredients
  - What is the Data Maturity Model? (DMM)
  - Body of Knowledge (DM BOK)
- Understanding and applying them together
  - Weakest link in the chain architecture
  - Just a bit on strategy
  - Three legged stool
  - How does one get to Carnegie Hall?
- Where to next?
- Q & A?

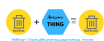
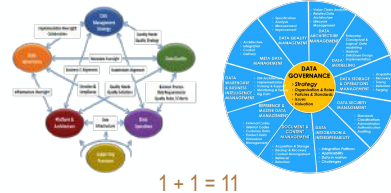
Practicing  
Data Management  
Better



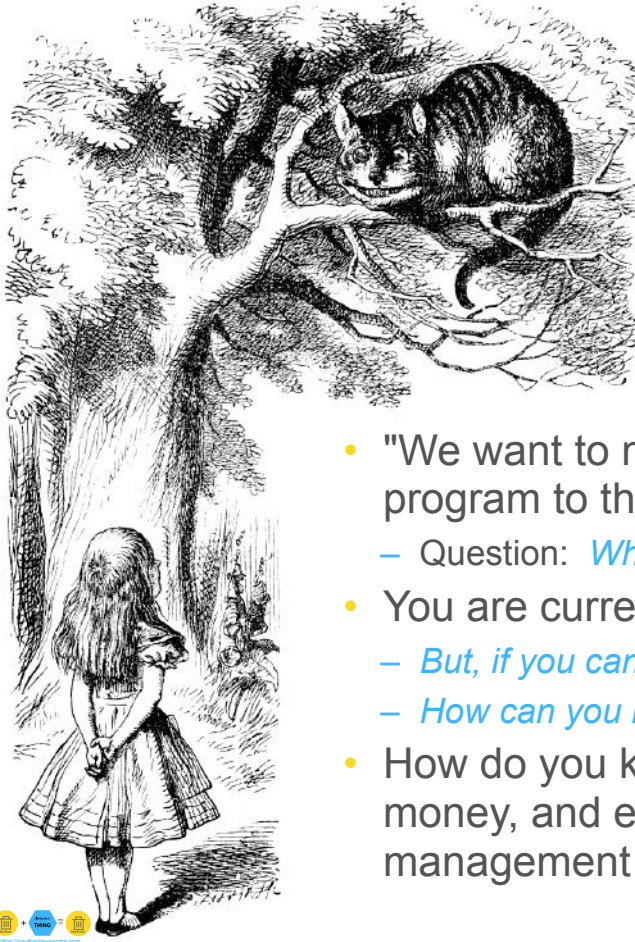
# Program overview

- Motivation
  - Frustration—we are unsatisfied with current state
  - Are we making progress? (No)
- How did we get here? (Building on proven research)
  - DoD → SEI → MITRE → CMMI
  - Industry push for best practices
- Ingredients
  - What is the Data Maturity Model? (DMM)
  - Body of Knowledge (DM BOK)
- Understanding and applying them together
  - Weakest link in the chain architecture
  - Just a bit on strategy
  - Three legged stool
  - How does one get to Carnegie Hall?
- Where to next?
- Q & A?

Practicing  
Data Management  
Better







## Motivation



"One day Alice came to a fork in the road and saw a Cheshire cat in a tree. Which road do I take? she asked. Where do you want to go? was his response. I don't know, Alice answered. Then, said the cat, it doesn't matter."

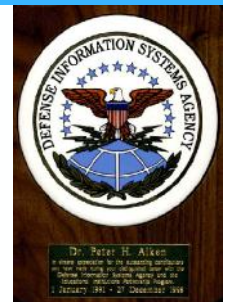
Lewis Carroll from *Alice in Wonderland*

- "We want to move our data management program to the next level"
  - Question: *What level are you at now?*
- You are currently managing your data,
  - *But, if you can't measure it,*
  - *How can you manage it effectively?*
- How do you know where to put time, money, and energy so that data management best supports the mission?

© Copyright 2023 by Peter Allen Slide 6 17

## DoD Origins

- US DoD Reverse Engineering Program Manager
- We sponsored research at the CMM/SEI asking
  - "How can we measure the performance of DoD and our partners?"
  - "Go check out what the Navy is up to!"
- SEI responded with an integrated process/data improvement approach
  - DoD required SEI to remove the data portion of the approach
  - It grew into CMMI/DM BoK, etc.



© Copyright 2023 by Peter Allen Slide 6 18

# Measuring Data Management Practice Maturity: A Community's Self-Assessment



*Peter Aiken*, Virginia Commonwealth University/Institute for Data Research  
*M. David Allen*, Data Blueprint  
*Burt Parker*, Independent consultant  
*Angela Mattia*, J. Sergeant Reynolds Community College

## MITRE Corporation: Data Management Maturity Model

- Internal research project: Oct '94-Sept '95
- Based on Software Engineering Institute Capability Maturity Model (SEI CMM<sup>SM</sup>) for Software Development Projects
- Key Process Areas (KPAs) parallel SEI CMM<sup>SM</sup> KPAs, but with data management focus and key practices
- Normative model for data management required; need to:
  - Understand scope of data management
  - Organize data management key practices
- Reported as not-done-well by those who do it

Increasing data management practice maturity levels can positively impact the coordination of data flow among organizations, individuals, and systems. Results from a self-assessment provide a roadmap for improving organizational data management practices.



© Copyright 2023 by Peter Aiken Slide 6 19

## Melanie Mecca

- Former CMMI Institute/Director of Data Management Products and Services → datawise.inc
- 30+ years designing and implementing strategies and solutions for private/public sectors
- Architecture/Design experience in:
  - Data Management Programs
  - Enterprise Data Architecture
  - Enterprise Architecture
- DMM's Managing Author  
Certified Partner, CMMI Institute

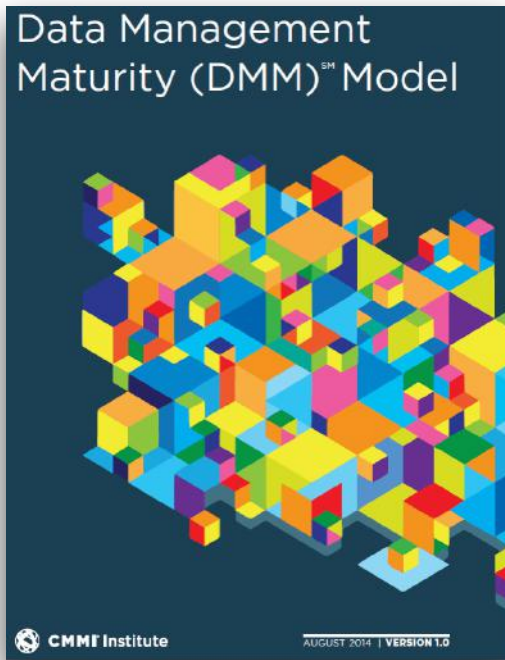


– [melanie@datawise-inc.com](mailto:melanie@datawise-inc.com)



© Copyright 2023 by Peter Aiken Slide 6 20

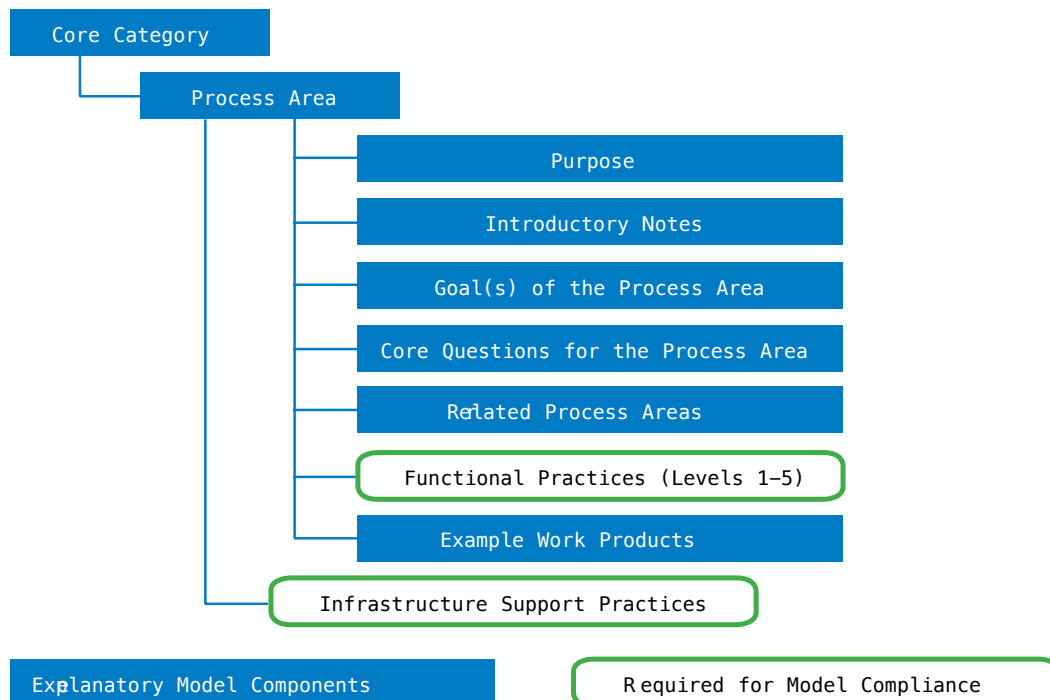
# Data Management Maturity (DMM)<sup>SM</sup> Model



- DMM 1.0 released August 2014
  - 3.5 years in development
  - Sponsors – Microsoft, Lockheed Martin, Booz Allen Hamilton
  - 50+ contributing authors, 70+ peer reviewers, 80+ orgs
- Reference model framework of fundamental best practices
  - 414 specific practice statements
  - 596 functional work products
  - Maturity practices
- Measurement instrument for organizations to evaluate capabilities and maturity, identify gaps, and incorporate guidelines for improvements.



## DMM Structure



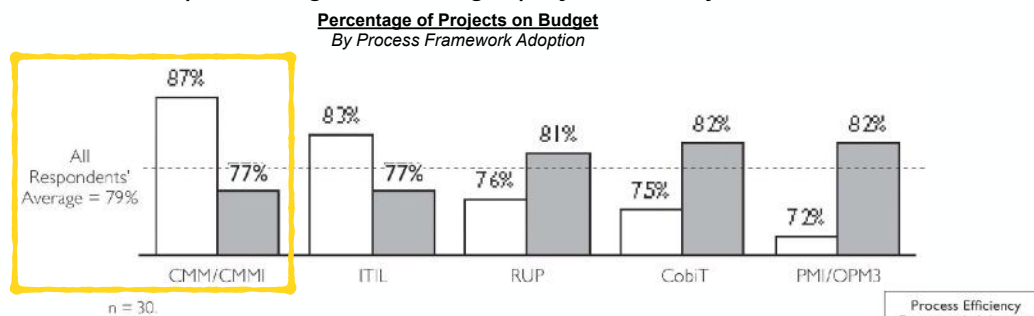
# “You Are What You DO”

- Model emphasizes behavior
  - Proactive positive behavioral changes
  - Creating and carrying out effective, repeatable processes
  - Leveraging and extending across the organization
- Activities result in work products
  - Processes, standards, guidelines, templates, policies, etc.
  - Reuse and extension = maximum value, lower costs, happier staff
- Practical focus reflects real-world organizations – enterprise program evolving to all hands on deck.

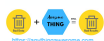
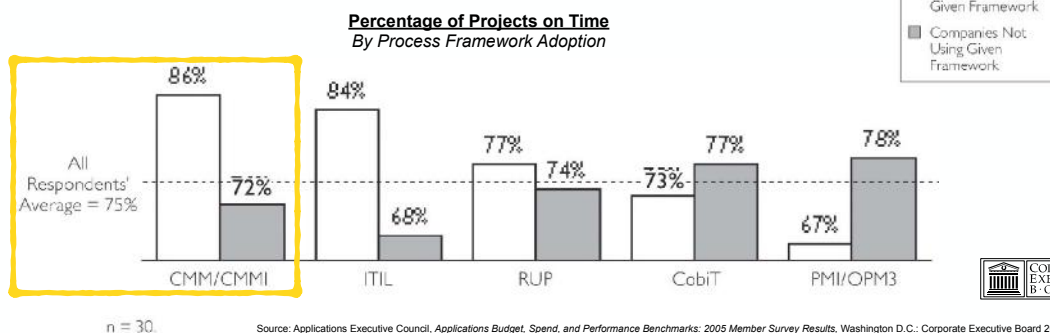


## Key Finding: Process Frameworks are not Created Equal

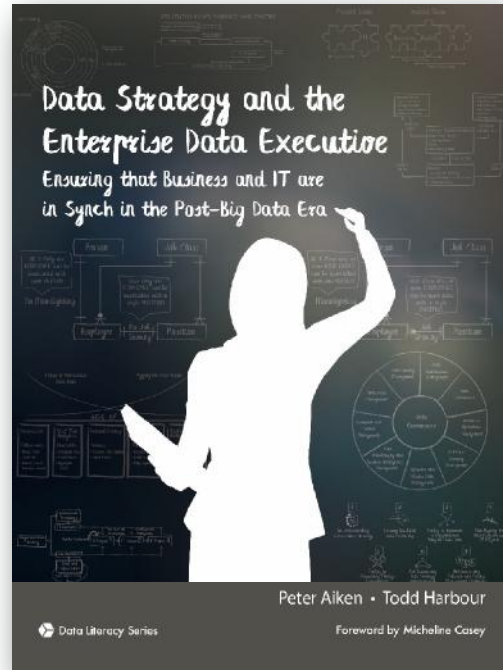
*With the exception of CMM and ITIL, use of process-efficiency frameworks does not predict higher on-budget project delivery...*



*...while the same pattern generally holds true for on-time performance*



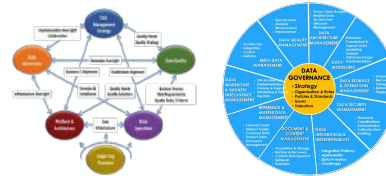
"While all improvement efforts begin with the obligatory 'assessment' phase, Carnegie Mellon's CMMI and DMM are the only proven frameworks that have the added benefit of literally decades of practice and benchmarking data. Organizations not using the DMM risk an inability to meaningfully compare results against other organizations and, as a result, adopt unproven methods."



## Program overview

- Motivation
  - Frustration—we are unsatisfied with current state
  - Are we making progress? (No)
- How did we get here? (Building on proven research)
  - DoD → SEI → MITRE → CMMI
  - Industry push for best practices
- Ingredients
  - What is the Data Maturity Model? (DMM)
  - Body of Knowledge (DM BOK)
- Understanding and applying them together
  - Weakest link in the chain architecture
  - Just a bit on strategy
  - Three legged stool
  - How does one get to Carnegie Hall?
- Where to next?
- Q & A?

## Practicing Data Management Better



1 + 1 = 11

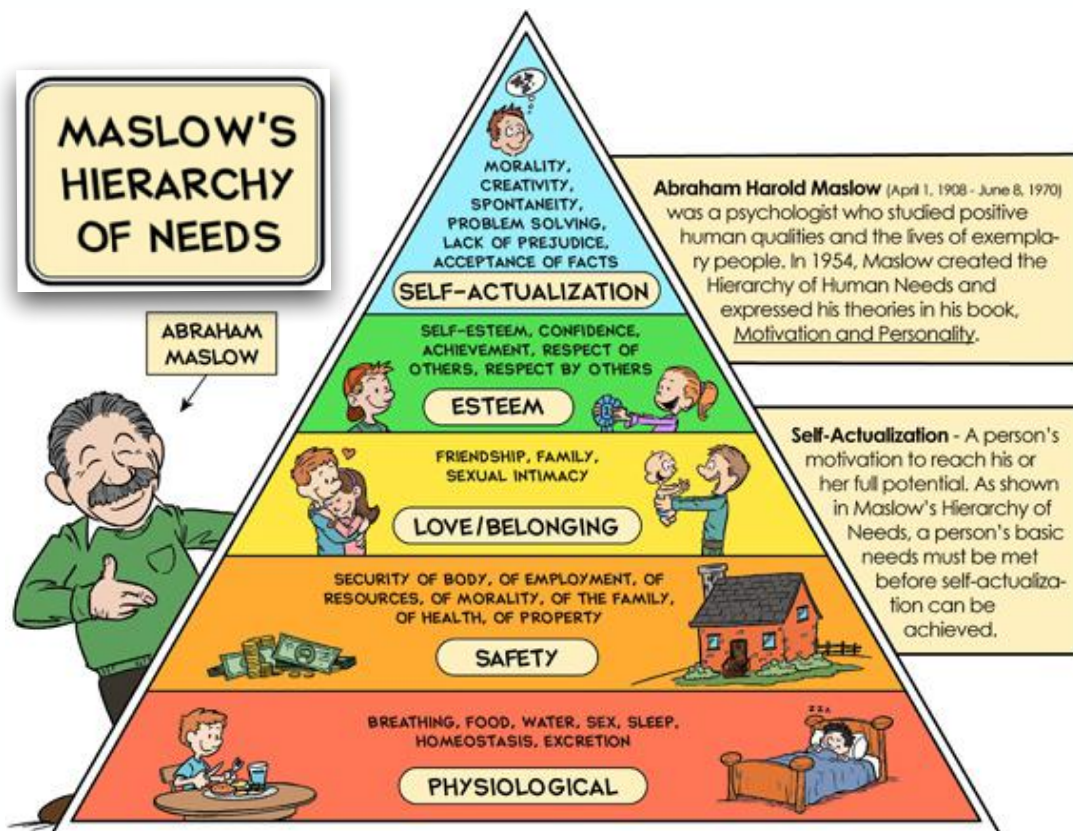


# Our barn had to pass a foundation inspection

- Before further construction could proceed
- It makes good business sense
- No IT equivalent



27

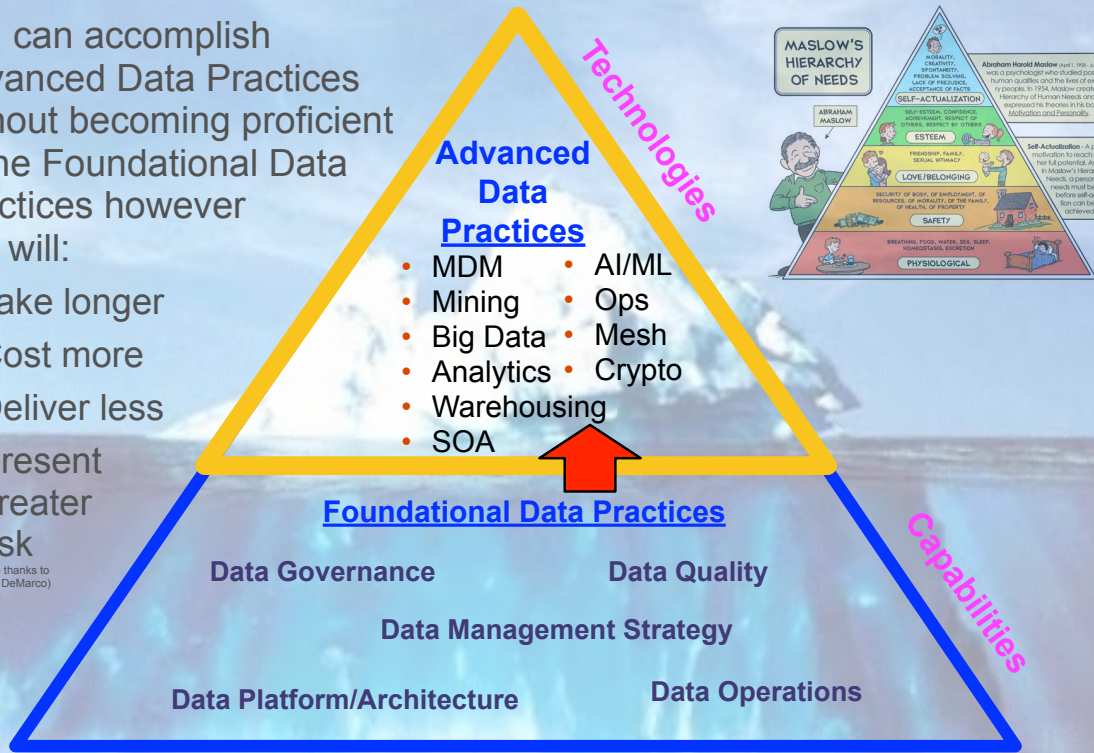


# Data Management Practices Hierarchy

You can accomplish Advanced Data Practices without becoming proficient in the Foundational Data Practices however this will:

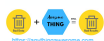
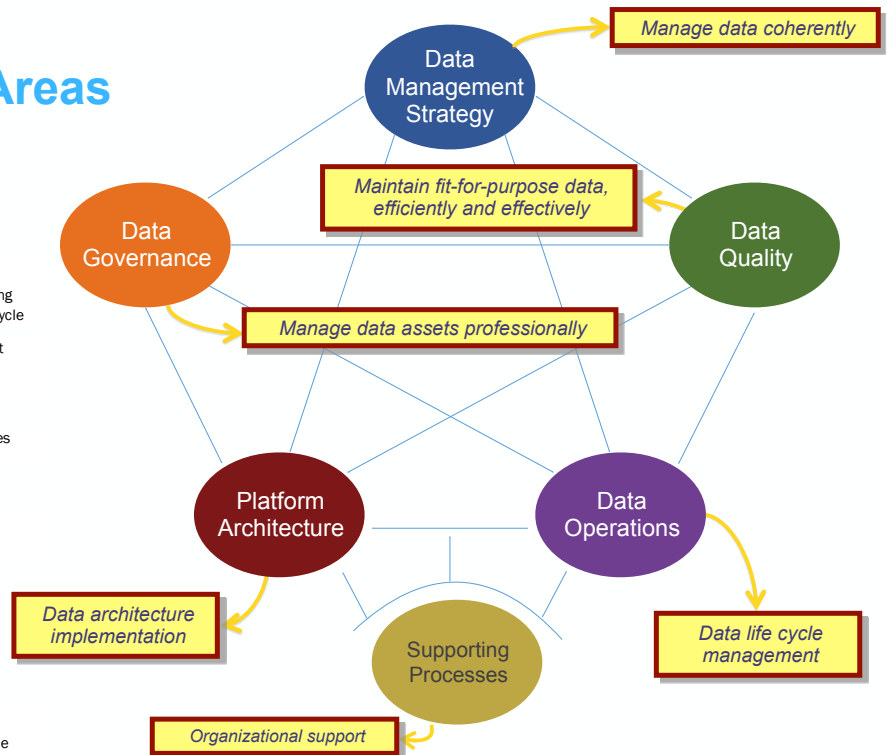
- Take longer
- Cost more
- Deliver less
- Present greater risk

(with thanks to Tom DeMarco)

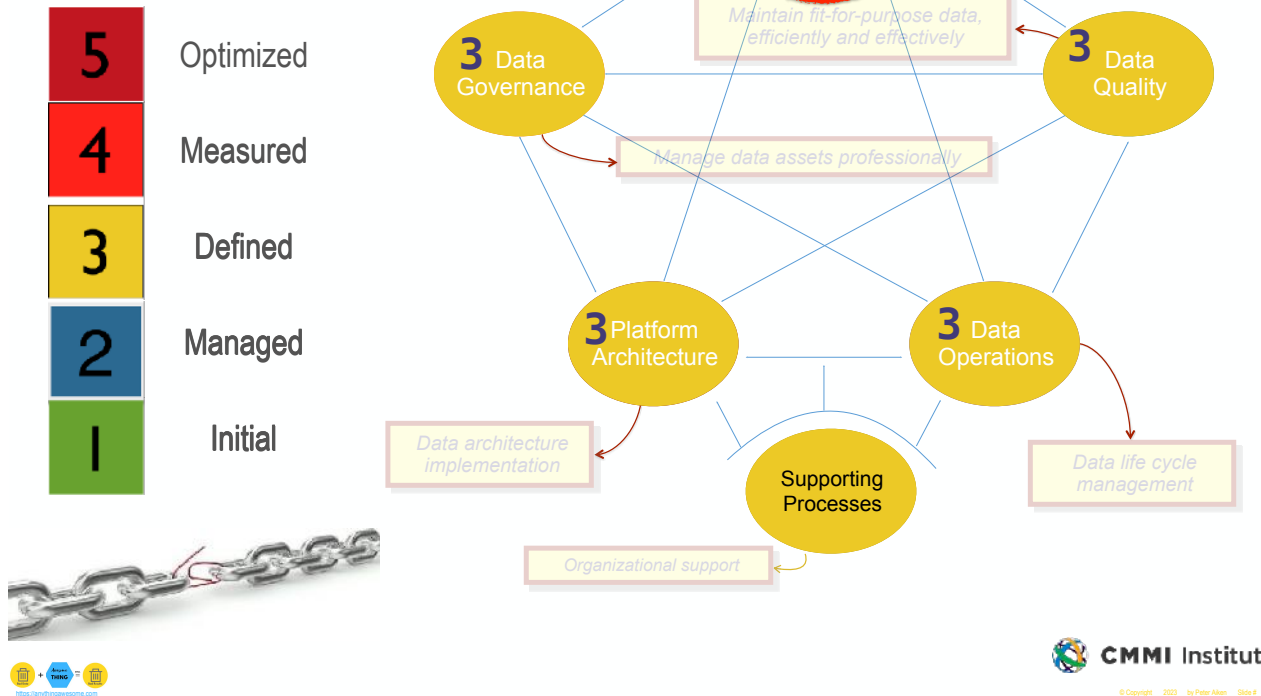


## DMM<sup>SM</sup> Structure of 5 Integrated DM Practice Areas

Component	Process Areas
Data Management Strategy	Data Management Goals Corporate Culture Data Management Funding Data Requirements Lifecycle
Data Governance	Governance Management Business Glossary Metadata Management
Data Operations	Standards and Procedures Data Sourcing
Data Quality	Data Quality Framework Data Quality Assurance
Platform & Architecture	Architectural Framework Platforms & Integration
Supporting Processes	Measurement & Analysis Process Management Process Quality Assurance Risk Management Configuration Management



# Your data foundation can only be as strong as its weakest link!



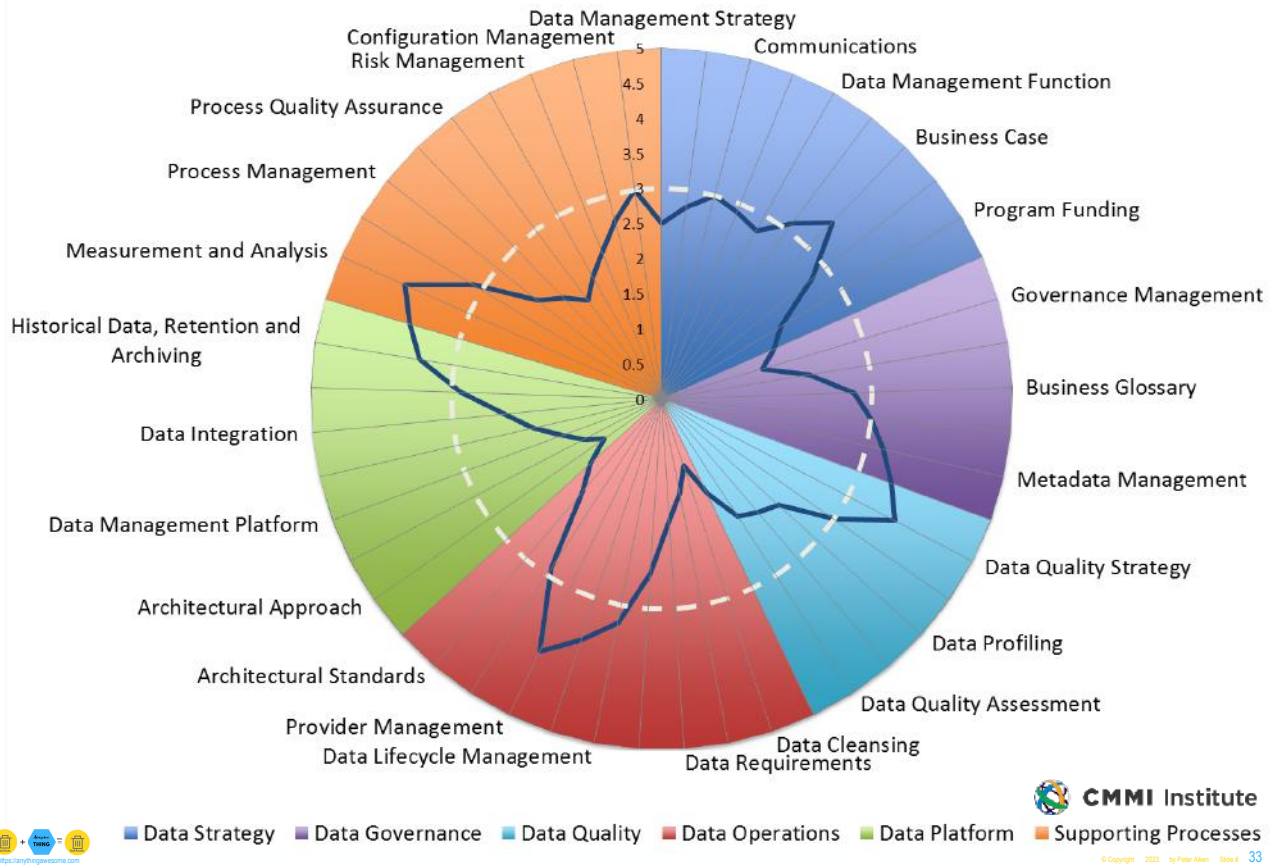
## Assessment Components



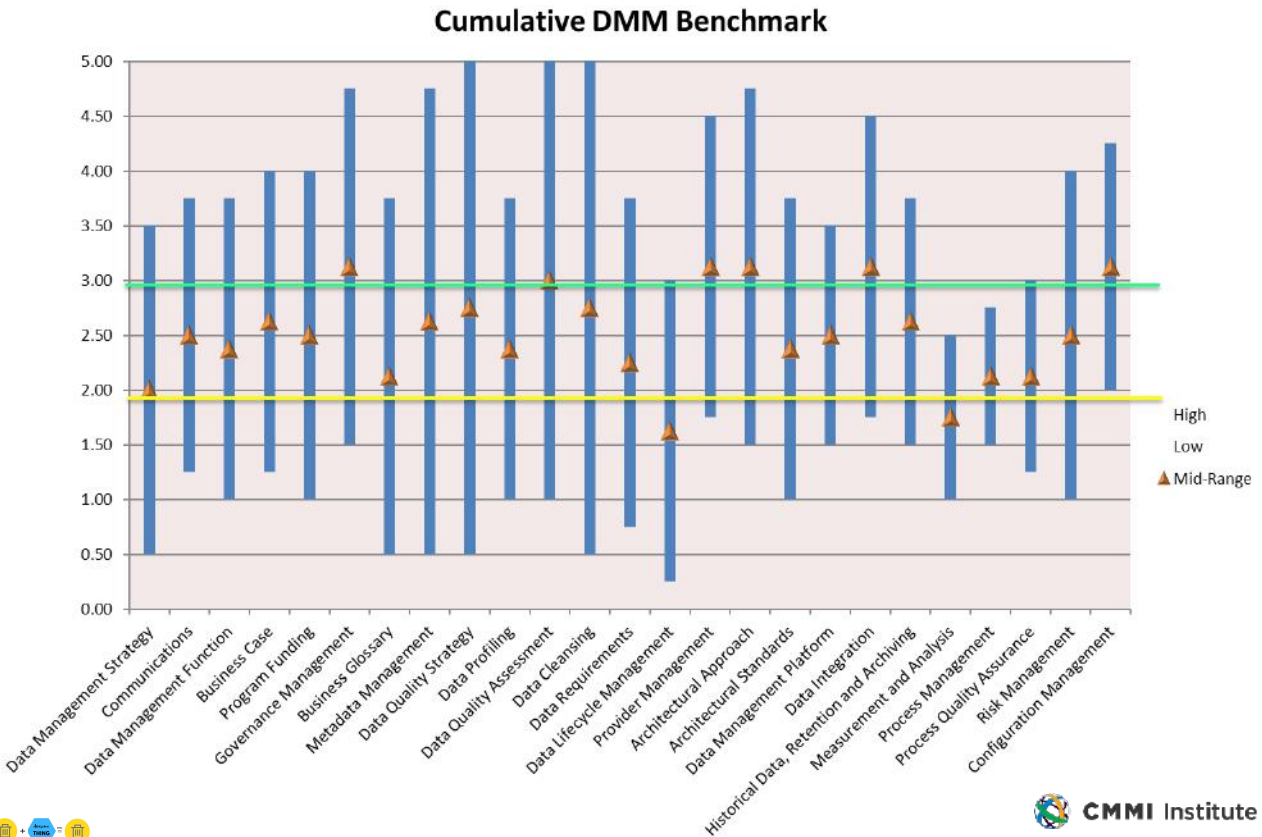
Data Management Practice Areas		Capability Maturity Model Levels	Examples of practice maturity
Data Management Strategy	DM is practiced as a coherent and coordinated set of activities	1 – Performed	Our DM practices are ad hoc and dependent upon "heroes" and heroic efforts
Data Quality	Delivery of data is support of organizational objectives – <i>the currency of DM</i>	2 – Managed	We have DM experience and have the ability to implement disciplined processes
Data Governance	Designating specific individuals caretakers for certain data	3 – Defined	We have standardized DM practices so that all in the organization can perform it with uniform quality
Data Platform/ Architecture	Efficient delivery of data via appropriate channels	4 – Measured	We manage our DM processes so that the whole organization can follow our standard DM guidance
Data Operations	Ensuring reliable access to data	5 – Optimized	We have a process for improving our DM capabilities



# Sample Assessment Summary



# Cumulative Benchmark – Multiple organizations



# Assessments

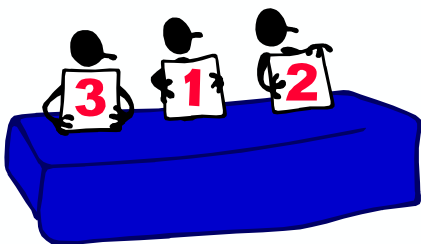
# ASSESSMENT

- It is generally not worth a lot of investment to discover that you are at the very beginning of your journey
- Use it to uncover previously unknown pockets of excellence
- First plan should be examine the feasibility of expanding these to other parts of the organization



# Industry Focused Results

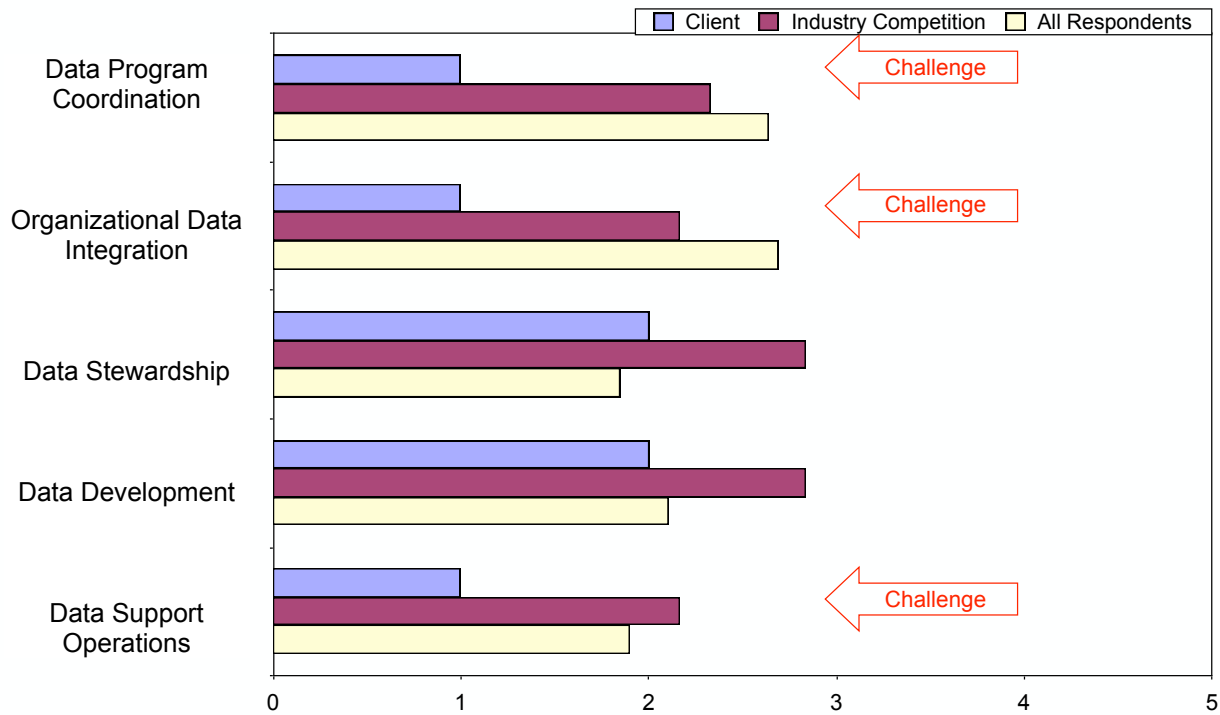
- CMU's Software Engineering Institute (SEI) Collaboration
- Results from hundreds organizations in various industries including:
  - ✓ Public Companies
  - ✓ State Government Agencies
  - ✓ Federal Government
  - ✓ International Organizations
- Defined industry standard
- Steps toward defining data management "state of the practice"



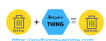
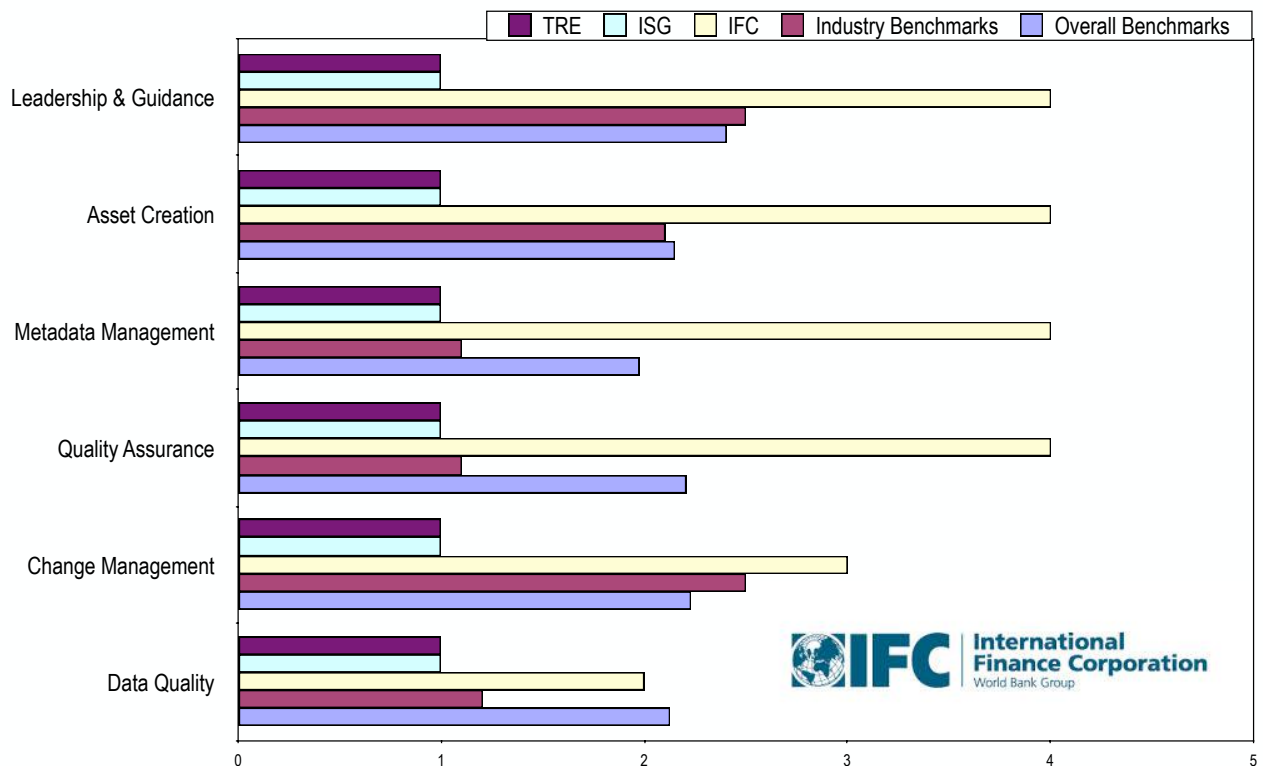
	Initial (I)	Managed (II)	Defined (III)	Measured (IV)	Optimized (V)
Data Management Strategy	Red	Yellow	Focus: Guidance and Facilitation		
Data Governance	Red	Yellow	Yellow	Yellow	Yellow
Platform & Architecture	Red	Yellow	Yellow	Yellow	Yellow
Data Quality	Red	Yellow	Focus: Implementation and Access		
Data Operations	Red	Yellow	Yellow	Yellow	Yellow



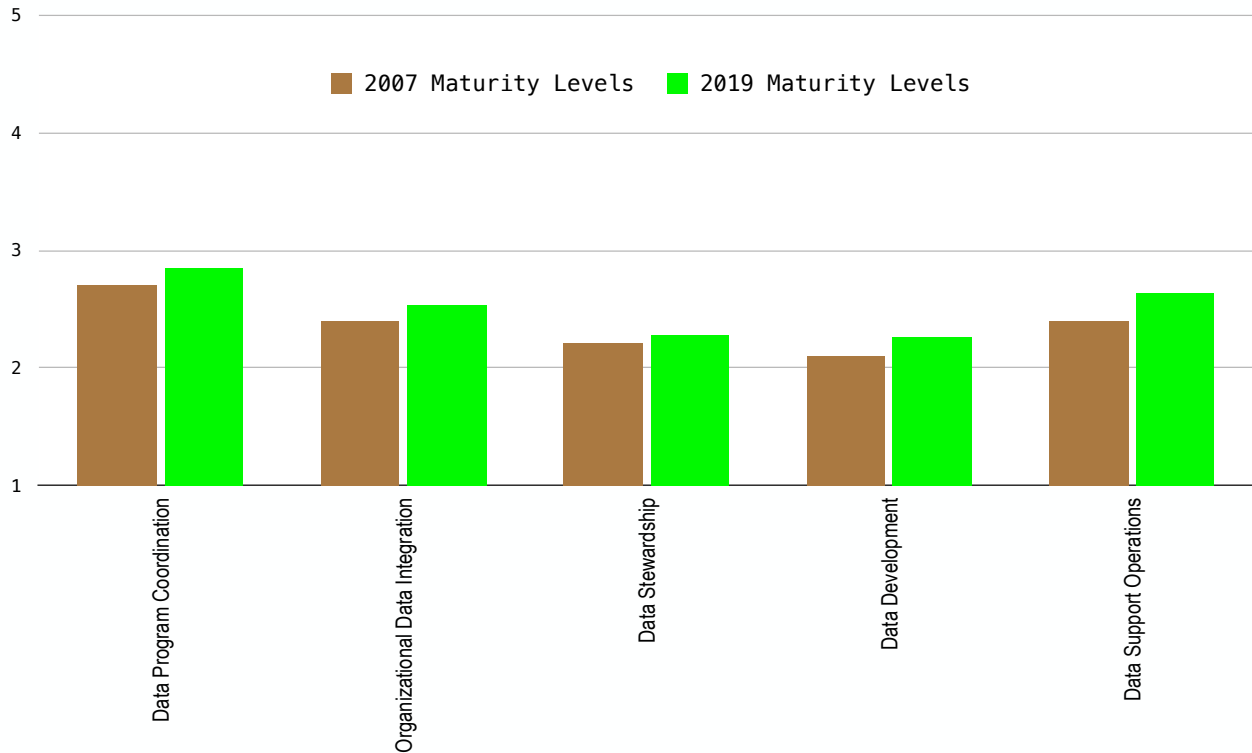
# Data Management Practices Assessment



# High Marks for IFC's Audit



# Comparison of DM Maturity 2007-2019



## How Literate are we?

### What is NAAL?

- a **N**ationally representative **A**ssessment of English **L**iteracy among American **A**dults age 16 and older NAAL → PIAAC (Program for the International Assessment of Adult Competencies)
- PIAAC assesses three key competencies for 21st-century society and the global economy



#### • **Literacy**

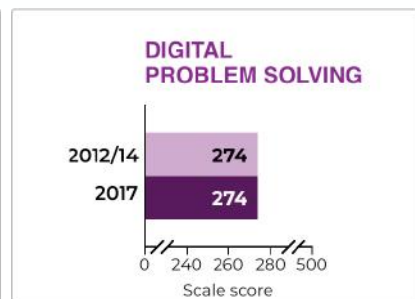
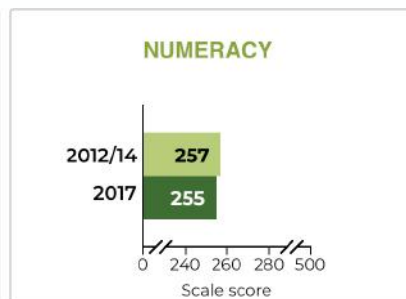
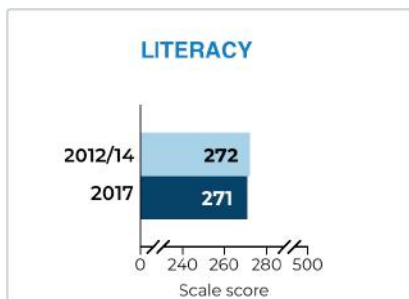
the ability to understand, use, and respond appropriately to written texts.

#### • **Numeracy**

the ability to use basic mathematical and computational skills.

#### • **Digital Problem Solving**

the ability to access/interpret information in digital environments to perform practical tasks.



[https://nces.ed.gov/surveys/piaac/current\\_results.asp](https://nces.ed.gov/surveys/piaac/current_results.asp)

- No statistically significant differences from 2012/14 to 2017!

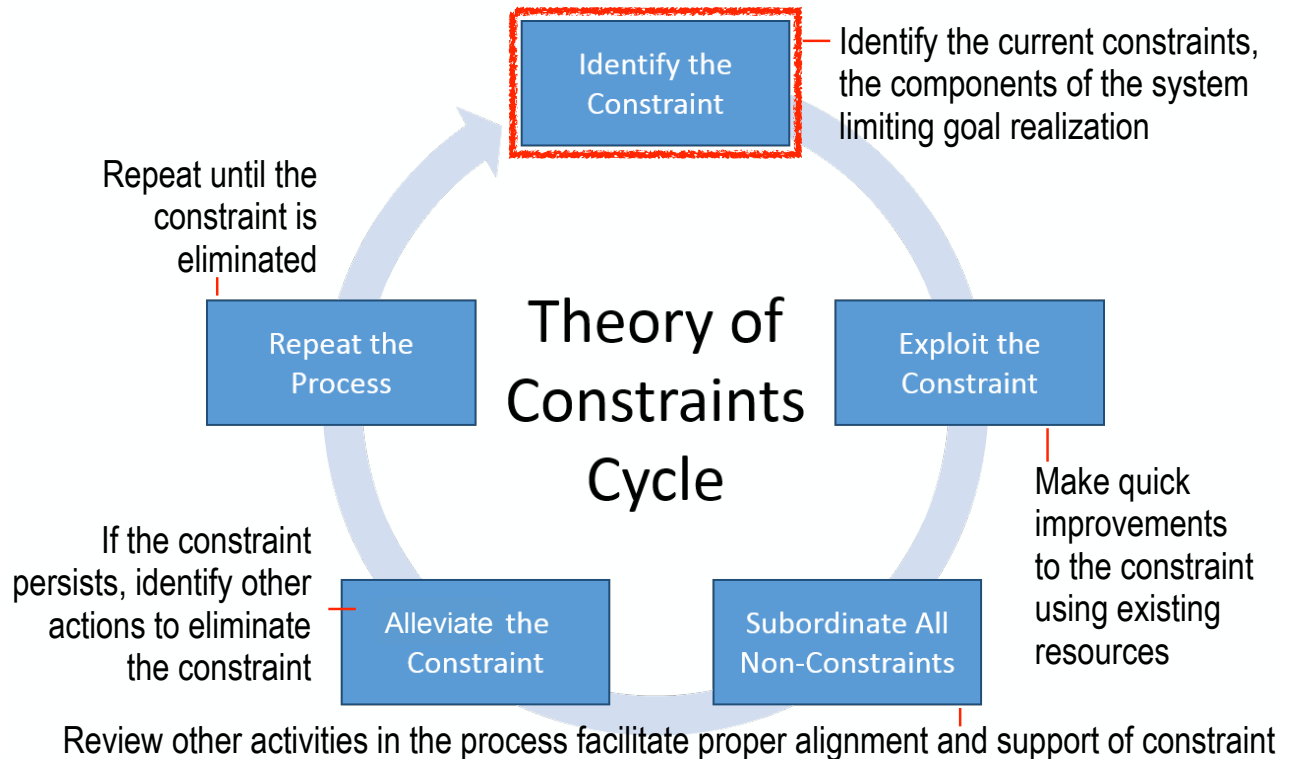


# A pattern in a stream of decisions

Mintzberg



## Theory of Constraints - Generic



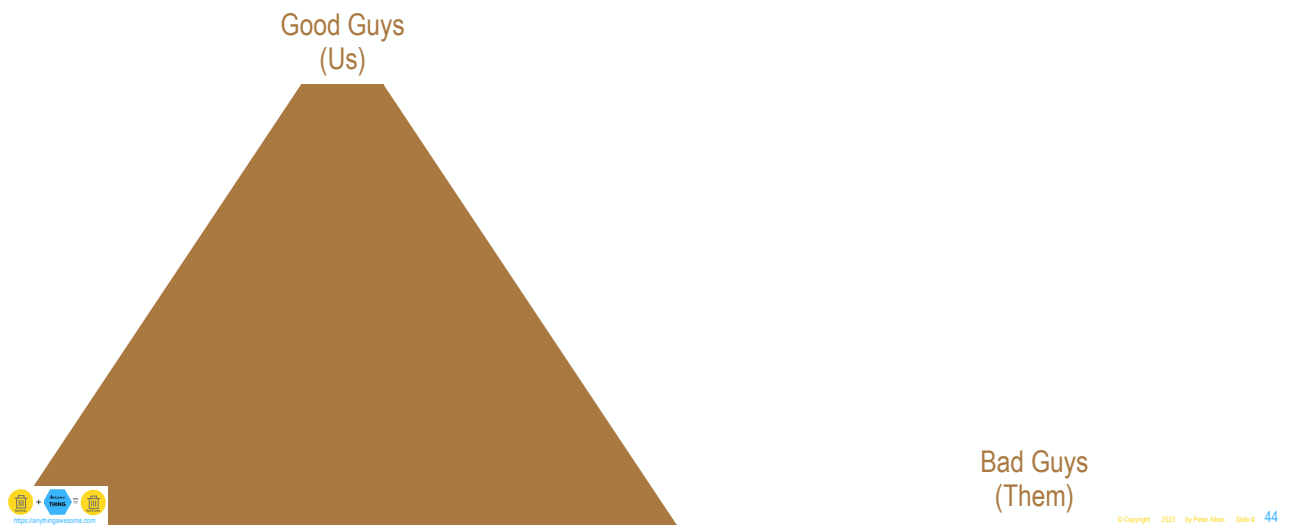
---

# Strategy Example 1

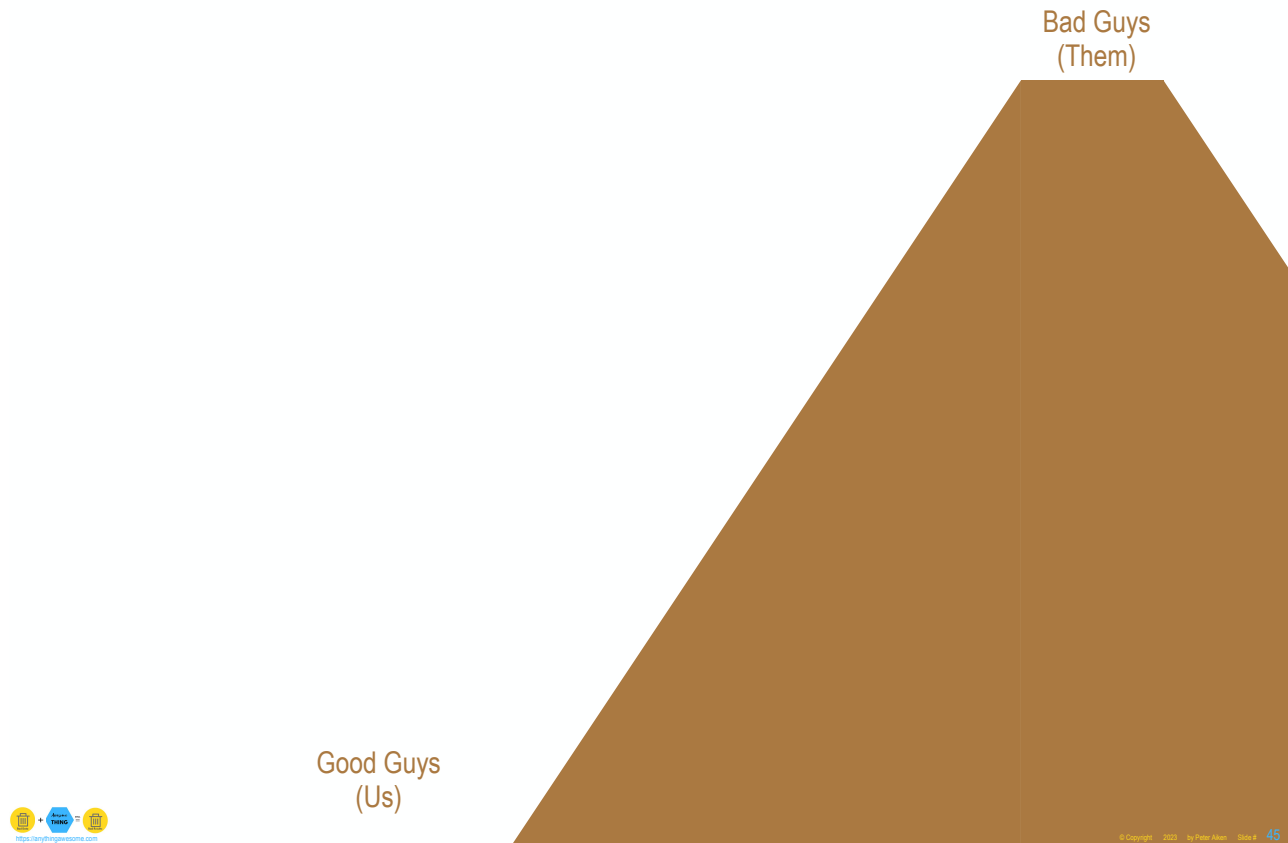


---

# Strategy Example 2



## Strategy Example 3



## General Dwight D. Eisenhower



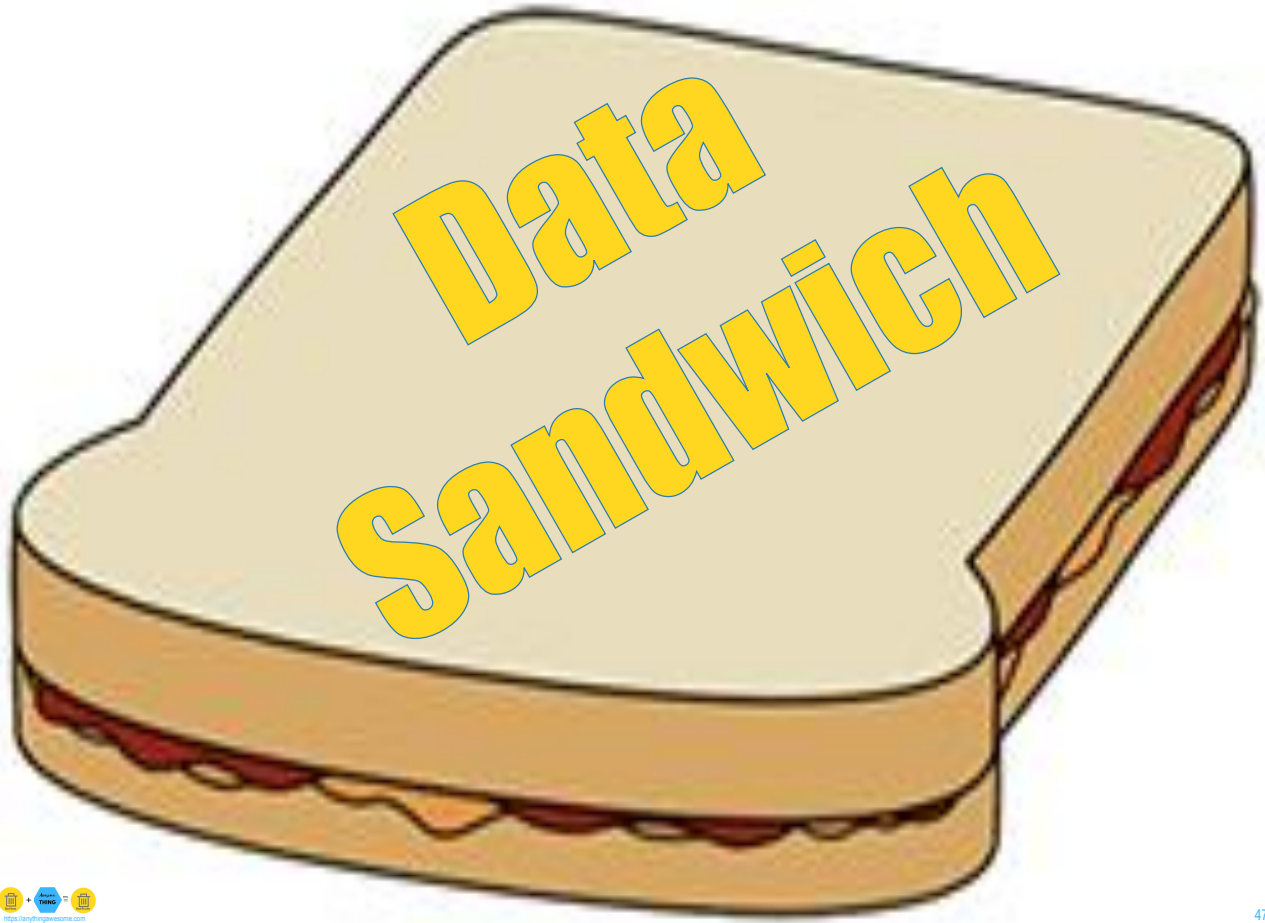
Military Plan



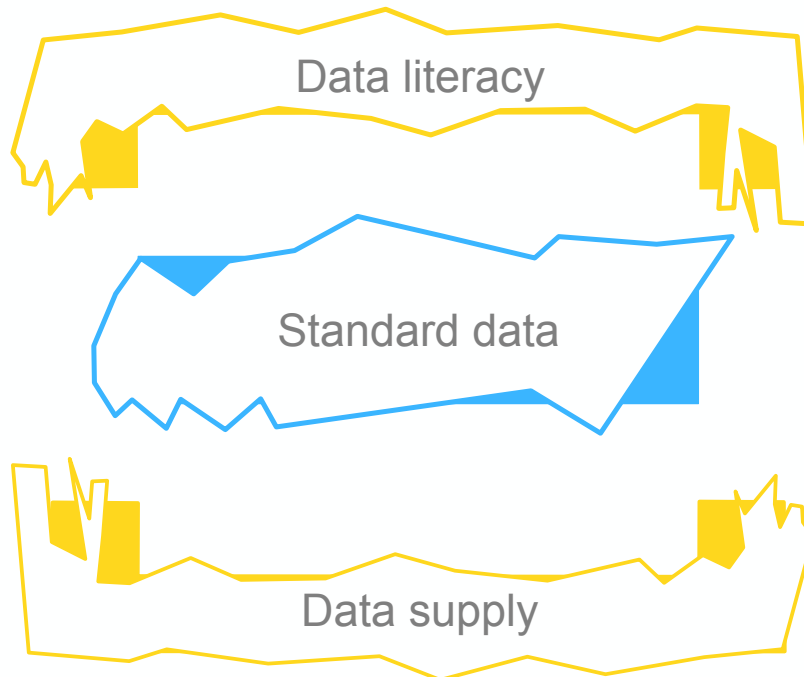
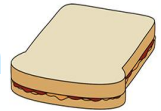
Dwight Eisenhower

**“In preparing for battle I have always found that plans are useless, but planning is indispensable ...”**

<https://quoteinvestigator.com/2017/11/18/planning/>

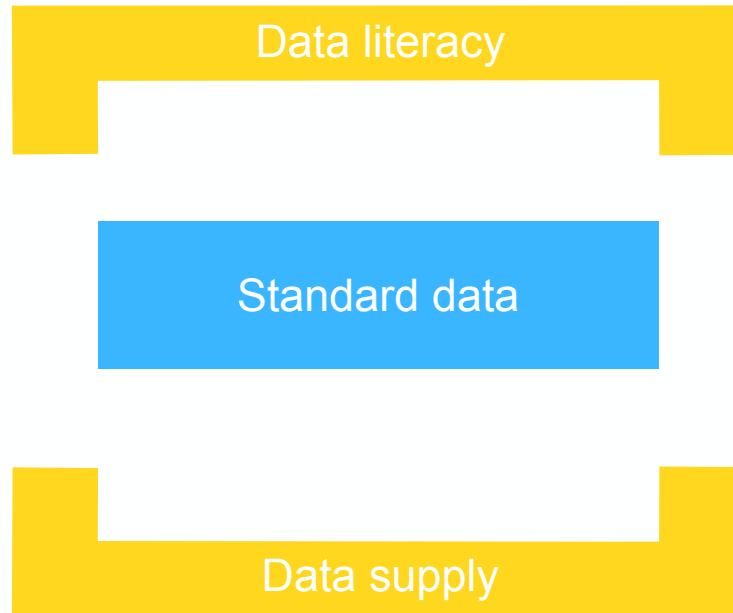
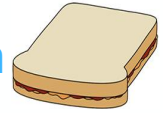


## Leverage point - high performance automation

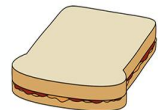




## Leverage point - high performance automation



## Leverage point - high performance automation

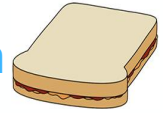


*This cannot happen without investments in engineering and architecture!*



**Quality engineering/  
architecture work products  
do not happen accidentally!**

# Leverage point - high performance automation

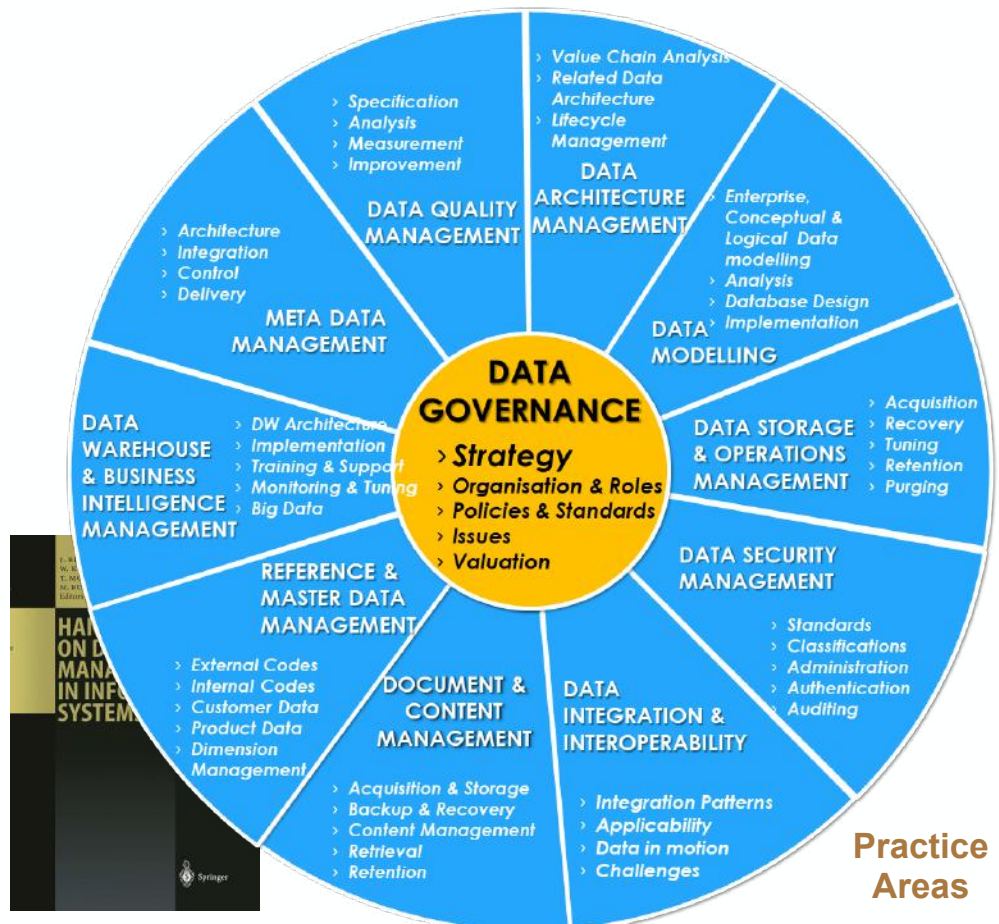


*This cannot happen without investments in data engineering and architecture!*



**Quality data engineering/  
architecture work products  
do not happen accidentally!**

## Data Management Body of Knowledge (DM BoK V2)

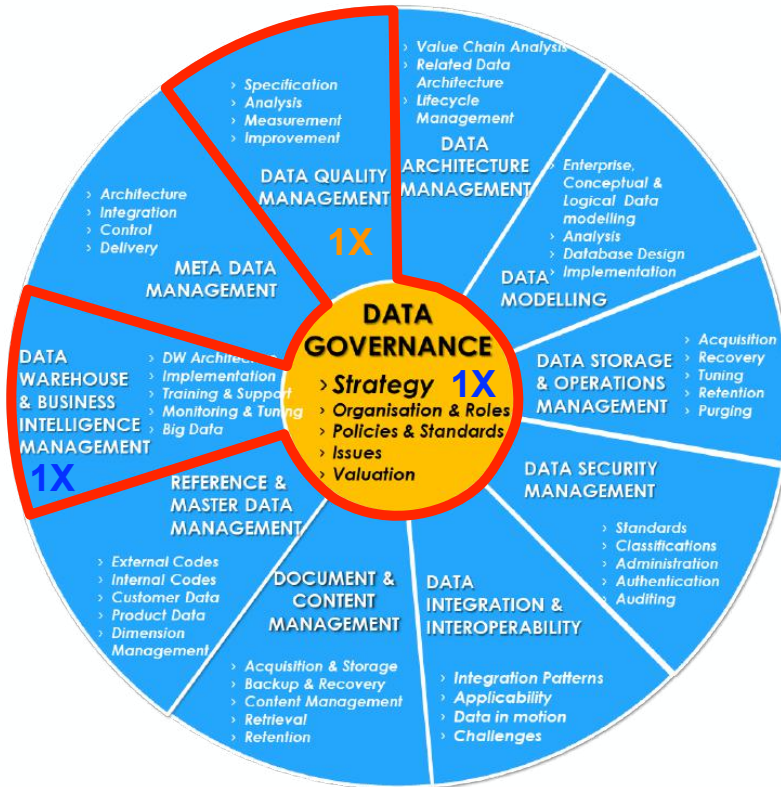
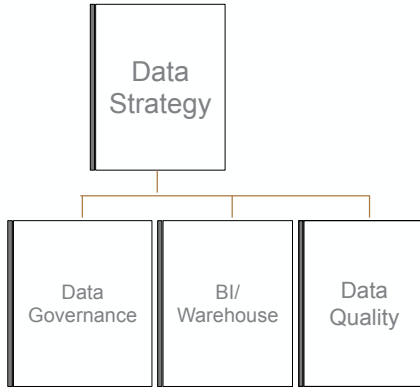


**Practice Areas**

# Version 1



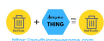
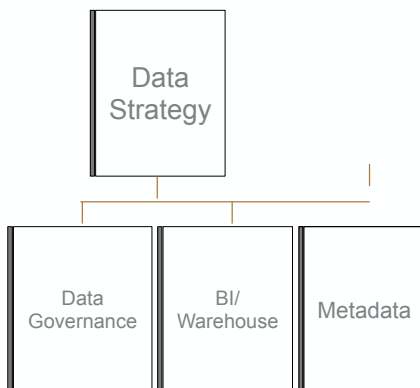
## Perfecting operations in 3 data management practice areas



# Version 2

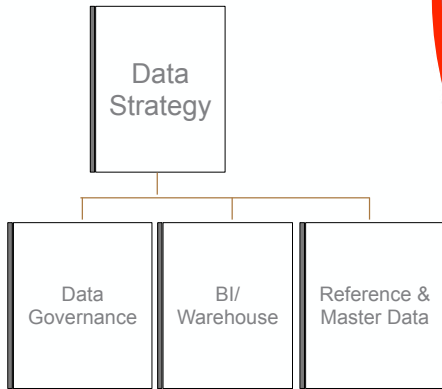


## Perfecting operations in 3 data management practice areas

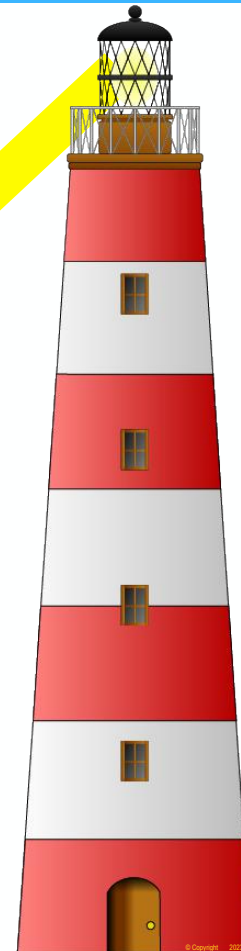




**Perfecting operations in 3 data management practice areas**



**Lighthouse Projects Provide Focus**





A Musical Analogy That Works for Both Practice and Storytelling

## Program Overview

- Motivation
  - Frustration—we are unsatisfied with current state
  - Are we making progress? (No)
- How did we get here? (Building on proven research)
  - DoD → SEI → MITRE → CMMI
  - Industry push for best practices
- Ingredients
  - What is the Data Maturity Model? (DMM)
  - Body of Knowledge (DM BOK)
- Understanding and applying them together
  - Weakest link in the chain architecture
  - Just a bit on strategy
  - Three legged stool
  - How does one get to Carnegie Hall?
- Where to next?
- Q & A?

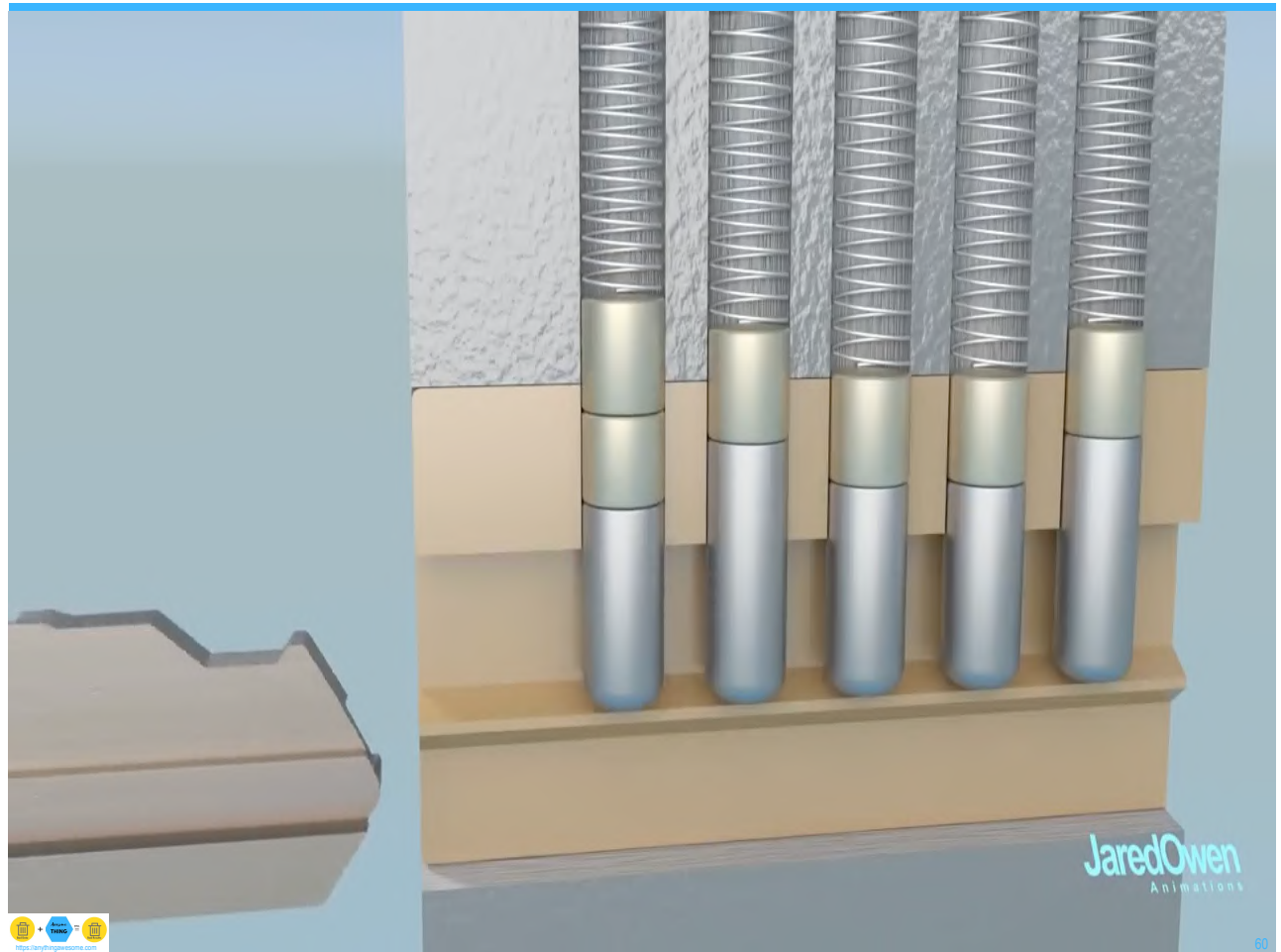
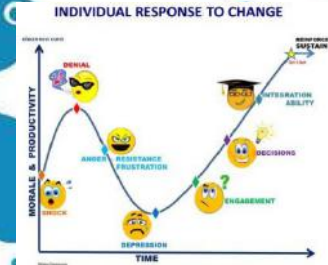
Practicing  
Data Management  
Better



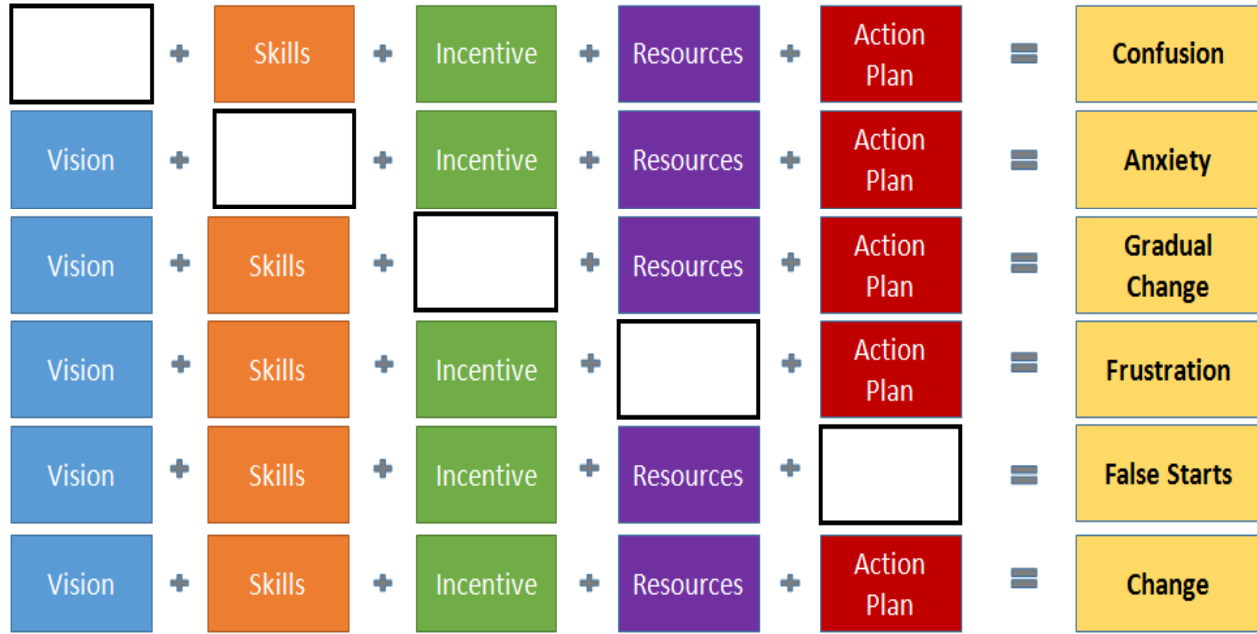
1 + 1 = 11



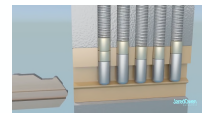
# Change Management & Leadership



# Diagnosing Organizational Readiness



**Culture is the biggest impediment to a shift in organizational thinking about data!**



adapted from the *Managing Complex Change* model by Lippitt, 1987

© Copyright 2023 by Peter Aiken Slide 61

## No cost, no registration case study download

ACM DIGITAL LIBRARY

Home > ACM Journals > Journal of Data and Information Quality (JDQ) > Vol. 7, No. 1-2 > EXPERIENCE: Succeeding at Data Management—BigCo Attempts to Leverage Data

RESEARCH-ARTICLE OPEN ACCESS

### EXPERIENCE: Succeeding at Data Management—BigCo Attempts to Leverage Data

Author: Peter Aiken [Authors Info & Affiliations](#)

Publication: Journal of Data and Information Quality (JDQ) • May 2016 • Article No. 8  
<https://doi.org/10.1145/2893482>

3,098

eRead PDF

- Download
- <http://dl.acm.org/citation.cfm?doid=2888577.2893482>

### EXPERIENCE: Succeeding at Data Management—BigCo Attempts to Leverage Data

PETER AIKEN, Virginia Commonwealth University/Data Blueprint

In a manner similar to most organizations, BigCompany (BigCo) was determined to benefit strategically from its widely recognized and vast quantities of data. (U.S. government agencies make regular visits to BigCo to learn from its experiences in this area.) When faced with an explosion in data volume, increases in complexity, and a need to respond to changing conditions, BigCo struggled to respond using a traditional, information technology (IT) project-based approach to address these challenges. As BigCo was not data knowledgeable, it did not realize that traditional approaches could not work. Two full years into the initiative, BigCo was far from achieving its initial goals. How much more time, money, and effort would be required before results were achieved? Moreover, could the results be achieved in time to support a larger, critical, technology-driven challenge that also depended on solving the data challenges? While these questions remain unanswered, these considerations increase our collective understanding of data assets as separate from IT projects. Only by reconceiving data as a strategic asset can organizations begin to address these new challenges. Transformation to a data-driven culture requires far more than technology, which remains just one of three required "steel legs" (people and process being the other two). Seven prerequisites to effectively leveraging data are necessary, but insufficient awareness exists in most organizations—hence, the widespread misfires in these areas, especially when attempting to implement the so-called big data initiatives. Refocusing on foundational data management practices is required for all organizations, regardless of their organizational or data strategies.

Categories and Subject Descriptors: H.2.0 (Information Systems): Database Management—General; E.0 (Data): General

General Terms: Management, Performance, Design

Additional Key Words and Phrases: Data management, data governance, data stewardship, organizational design, CDO, CIO, chief data officer, chief information officer, data, data architecture, enterprise data executive, IT management, strategy, policy, enterprise architecture, information systems, conceptual modeling, data integration, data warehousing, analytics, and business intelligence, BigCo

#### ACM Reference Format:

Peter Aiken. 2016. Experience: Succeeding at data management—BigCo attempts to leverage data. *J. Data and Information Quality* 7, 1-2, Article 8 (May 2016), 35 pages.  
 DOI: <http://dx.doi.org/10.1145/2893482>

#### 1. CASE INTRODUCTION

*Good technology in the hands of an inexperienced user rarely produces positive results.*

Everyone wants to "leverage" data. Today, this is most often interpreted as investments in warehousing, analytics, business intelligence (BI), and so on. After all, that is what you do with an asset—you leverage it—so the asset can help you to attain strategic objectives; see Redman [2008] and Ladley [2010]. Widespread and pervasive

Author's address: P. Aiken, 10124C West Broad Street, Glen Allen, VA 23060; email: [peteraiken@vcu.edu](mailto:peteraiken@vcu.edu). Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions.acm.org](http://Permissions.acm.org).  
 2016 Copyright is held by the owner/authors. Publication rights licensed to ACM.  
 ACM 1066-1055/2016/05-ARTS 35:00  
 DOI: <http://dx.doi.org/10.1145/2893482>

ACM Journal of Data and Information Quality, Vol. 7, No. 1-2, Article 8. Publication date: May 2016.



© Copyright 2023 by Peter Aiken Slide 62

## Big changes

1. Data volume is still increasing faster than we are able to process it,
2. Data interchange overhead and other costs of poor data practices are measurably sapping organization and individual resources—and therefore productivity,
3. Reliance on existing technology-based approaches and education methods has not materially addressed this gap between creation and processing or reduced bottom line costs, &
4. There exists an industry-type, whose sole purpose is to extract data from citizens and then use it for to make money.



1. Process is more important than results at first
2. Failure is itself a lesson
3. People and process aspects are not receiving enough attention
4. Best practices do exist



## Upcoming Events

Time: 19:00 UTC (2:00 PM NYC) | Presented by: Peter Aiken, PhD

Data Strategy Best Practices  
9 January 2024



Data Modeling Fundamentals  
13 February 2023

The Roles of Data Stewards  
12 March 2023

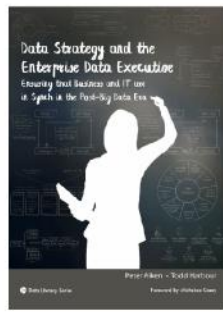
Brought to you by:





# Event Pricing

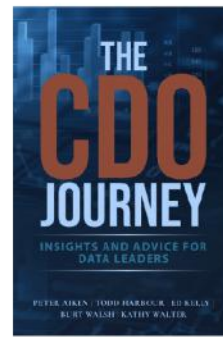
- 20% off directly from the publisher on select titles
- [My Book Store @ http://plusanythingawesome.com](http://plusanythingawesome.com)
- Enter the code "anythingawesome" at the Technics bookstore checkout where it says to "Apply Coupon"



**Data Strategy and the Enterprise Data Executive**

Ensuring that Business and IT are in Sync in the Post-Big Data Era

[Learn More](#)



**The CDO Journey**

Insights and Advice for Data Leaders

Note: the authors have donated all proceeds from this title towards the [International Society for Chief Data Officers](#).

[Learn More](#)



**Monetizing Data Management**

17 Case Studies Illustrating How Data Leveraging (Big and Small) Can Produce Quantifiable Results That Are of Keen Interest to C-Suite Occupants

[Learn More](#)

Coupon code

Apply coupon



Critical Design Review?

Mentoring?

Executive Data Literacy Training?

Collaboration?



[Peter.Aiken@AnythingAwesome.com](mailto:Peter.Aiken@AnythingAwesome.com) +1.804.382.5957



Independent Verification & Validation

Reverse Engineering Expertise?

Hiring Assistance?

# Thank You!

Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - <https://anythingawesome.com/OfficeHours.html>



# Measuring Data Management Practice Maturity: A Community's Self-Assessment

*Peter Aiken*, Virginia Commonwealth University/Institute for Data Research

*M. David Allen*, Data Blueprint

*Burt Parker*, Independent consultant

*Angela Mattia*, J. Sergeant Reynolds Community College

**Increasing data management practice maturity levels can positively impact the coordination of data flow among organizations, individuals, and systems. Results from a self-assessment provide a roadmap for improving organizational data management practices.**

**A**s increasing amounts of data flow within and between organizations, the problems that can result from poor data management practices are becoming more apparent. Studies have shown that such poor practices are widespread. For example,

- PricewaterhouseCoopers reported that in 2004, only one in three organizations were highly confident in their own data, and only 18 percent were very confident in data received from other organizations. Further, just two in five companies have a documented board-approved data strategy ([www.pwc.com/extweb/pwcpublishings.nsf/docid/15383D6E748A727DCA2571B6002F6EE9](http://www.pwc.com/extweb/pwcpublishings.nsf/docid/15383D6E748A727DCA2571B6002F6EE9)).
- Michael Blaha<sup>1</sup> and others in the research community have cited past organizational data management education and practices as the cause for poor database design being the norm.
- According to industry pioneer John Zachman,<sup>2</sup> organizations typically spend between 20 and 40 percent of their information technology budgets evolving their data via migration (changing data locations), con-

version (changing data into other forms, states, or products), or scrubbing (inspecting and manipulating, recoding, or rekeying data to prepare it for subsequent use).

- Approximately two-thirds of organizational data managers have formal data management training; slightly more than two-thirds of organizations use or plan to apply formal metadata management techniques; and slightly fewer than one-half manage their metadata using computer-aided software engineering tools and repository technologies.<sup>3</sup>

When combined with our personal observations, these results suggest that most organizations can benefit from the application of organization-wide data management practices. Failure to manage data as an enterprise-, corporate-, or organization-wide asset is costly in terms of market share, profit, strategic opportunity, stock price, and so on. To the extent that world-class organizations have shown that opportunities can be created through the effective use of data, investing in data as the only organizational asset that can't be depleted should be of great interest.

**Table 1. Data management processes.**<sup>4</sup>

Process	Description	Focus	Data type
Data program coordination	Provide appropriate data management process and technological infrastructure	Direction	Program data: Descriptive propositions or observations needed to establish, document, sustain, control, and improve organizational data-oriented activities (such as vision, goals, policies, and metrics).
Organizational data integration	Achieve organizational sharing of appropriate data	Direction	Development data: Descriptive facts, propositions, or observations used to develop and document the structures and interrelationships of data (for example, data models, database designs, and specifications).
Data stewardship	Achieve business-entity subject area data integration	Direction and implementation	Stewardship data: Descriptive facts about data documenting semantics and syntax (such as name, definition, and format).
Data development	Achieve data sharing within a business area	Implementation	Business data: Facts and their constructs used to accomplish enterprise business activities (such as data elements, records, and files).
Data support operations	Provide reliable access to data	Implementation	
Data asset use	Leverage data in business activities	Implementation	

## DATA MANAGEMENT DEFINITION AND EVOLUTION

As Table 1 shows, data management consists of six interrelated and coordinated processes, primarily derived by Burt Parker from sponsored research he led for the US Department of Defense at the MITRE Corporation.<sup>4</sup>

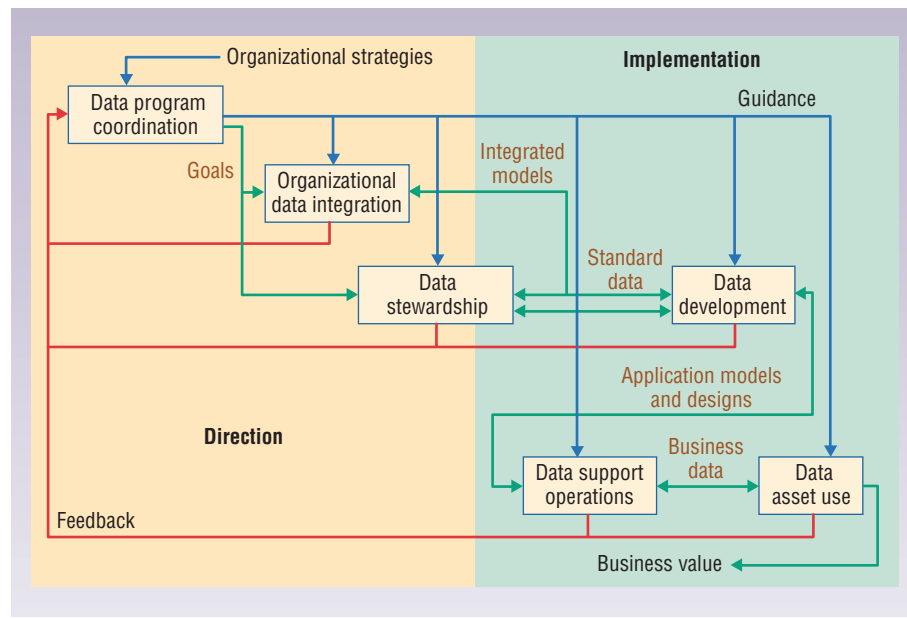
Figure 1 supports the similarly standardized definition: “Enterprise-wide management of data is understanding the current and future data needs of an enterprise and making that data effective and efficient in supporting business activities.”<sup>4</sup>

The figure illustrates how organizational strategies guide other data management processes. Two of these processes—data program coordination and organizational data integration—provide direction to the implementation processes—data development, data support operations, and data asset use. The data stewardship process straddles the line between direction and implementation. All processes exchange feedback designed to improve and fine-tune overall data management practices.

Data management has existed in some form since the 1950s and has been recognized as a discipline since the 1970s. Data management is thus a young discipline compared to, for example, the relatively mature

accounting practices that have been practiced for thousands of years. As Figure 2 shows, data management’s scope has expanded over time, and this expansion continues today.

Ideally, organizations derive their data management requirements from enterprise-wide information and functional user requirements. Some of these requirements come from legacy systems and off-the-shelf software packages. An organization derives its future data requirements from an analysis of what it will deliver, as well as future capabilities it will need to implement organizational strategies. Data management guides the trans-



**Figure 1. Interrelationships among data management processes (adapted from Burt Parker’s earlier work<sup>4</sup>). Blue lines indicate guidance, red lines indicate feedback, and green lines indicate data.**

Expanding Data Management Scope	1950-1970	1970-1990	1990-2000	2000 to present
Database development Database operation				
Data requirements analysis Data modeling				
Enterprise data management coordination Enterprise data integration Enterprise data stewardship Enterprise data use				
Explicit focus on data quality throughout Security Compliance Other responsibilities				

**Figure 2. Data management's growth over time. The discipline has expanded from an initial focus on database development and operation in the 1950s to 1970s to include additional responsibilities in the periods 1970-1990, 1990-2000, and from 2000 to the present.**

formation of strategic organizational information needs into specific data requirements associated with particular technology system development projects.

All organizations have data architectures, whether explicitly documented or implicitly assumed. An important data management process is to document the architecture's capabilities, making it more useful to the organization.

In addition, data management

- must be viewed as a means to an end, not the end itself. Organizations must not practice data management as an abstract discipline, but as a process supporting specific enterprise objectives—in particular, to provide a shared-resource basis on which to build additional services.
- involves both process and policy. Data management tasks range from strategic data planning to the creation of data element standards to database design, implementation, and maintenance.
- has a technical component: interfacing with and facilitating interaction between software and hardware.
- has a specific focus: creating and maintaining data to provide useful information.
- includes management of metadata artifacts that address the data's form as well as its content.

Although data management serves the organization, the organization often doesn't appreciate the value it provides. Some data management staffs keep ahead of the layoff curve by demonstrating positive business value. Management's short-term focus has often made it difficult to secure funding for medium- and long-term data management investments. Tracing the discipline's efforts to direct and indirect organizational benefits has been difficult, so it hasn't been easy to present an articulate business case to management that justifies subse-

quent strategic investments in data management.

Viewing data management as a collection of processes, each with a role that provides value to the organization through data, makes it easier to trace value through those processes and point not only to a methodological "why" of data management practice improvement but also to a specific, concrete "how."

## RESEARCH BASIS

Mark Gillenson has published three papers that serve as an excellent background to this research.<sup>5-7</sup> Like earlier works, Gillenson focuses on the implementation half of Figure 1, adopting a more narrow definition of

data administration. Over time, his work paints a picture of an industry attempting to catch up with technological implementation. Our work here updates and confirms his basic conclusions while changing the focus from whether a process is performed to the maturity with which it is performed.

Three other works also influenced our research: Ralph Keeney's value-focused thinking,<sup>8</sup> Richard Nolan's six-stage theory of data processing,<sup>9</sup> and the Capability Maturity Model Integration (CMMI).<sup>10,11</sup>

Keeney's value-focused thinking provides a methodological approach to analyzing and evaluating the various aspects of data management and their associated key process areas. We wove the concepts behind means and fundamental objectives into our assessment's construction to connect how we measure data management with what customers require from it.

In Stage VI of his six-stage theory of data processing, Nolan defined maturity as data resource management. Although Nolan's theory predates and is similar to the CMMI, it contains several ideas that we adapted and reused in the larger data management context. However, CMMI refinement remains our primary influence.

Most technologists are familiar with the CMM (and its upgrade to the CMMI), developed at Carnegie Mellon's Software Engineering Institute with assistance from the MITRE Corporation.<sup>10,11</sup> The CMMI itself was derived from work that Ron Radice and Watts Humphrey performed while at IBM. Dennis Goldenson and Diane Gibson presented results pointing to a link between CMMI process maturity and organizational success.<sup>12</sup> In addition, Cyndy Billings and Jeanie Clifton demonstrated the long-term effects for organizations that successfully sustain process improvement for more than a decade.<sup>13</sup>

CMMI-based maturity models exist for human resources, security, training, and several other areas of the software-related development process. Our colleague,

Brett Champlin, contributed a list of dozens of maturity measurements derived from or influenced by the CMMI. This list includes maturity measurement frameworks for data warehousing, metadata management, and software systems deployment. The CMMI's successful adoption in other areas encouraged us to use it as the basis for our data management practice assessment.

Whereas the core ideas behind the CMMI present a reasonable base for data management practice maturity measurement, we can avoid some potential pitfalls by learning from the revisions and later work done with the CMMI. Examples of such improvements include general changes to how the CMMI makes interrelationships between process areas more explicit and how it presents results to a target organization.

Work by Cynthia Hauer<sup>14</sup> and Walter Schnider and Klaus Schwinn<sup>15</sup> also influenced our general approach to a data management maturity model. Hauer nicely articulated some examples of the value determination factors and results criteria that we have adopted. Schnider and Schwinn presented a rough but inspirational outline of what mature data management practices might look like and the accompanying motivations.

## RESEARCH OBJECTIVES

Our research had six specific objectives, which we grouped into two types: *community descriptive goals* and *self-improvement goals*.

Community descriptive research goals help clarify our understanding of the data management community and associated practices. Specifically, we want to understand

- the range of practices within the data management community;
- the distribution of data management practices, specifically the various stages of organizational data management maturity; and
- the current state of data management practices—in what areas are the community data management practices weak, average, and strong?

Self-improvement research goals help the community as a whole improve its collective data management practices. Here, we desire to

- better understand what defines current data management practices;
- determine how the assessment informs our standing as a technical community (specifically, how does data management compare to software development?); and
- gain information useful for developing a roadmap for improving current practice.

The CMMI's stated goals are almost identical to ours: "[The CMMI] was designed to help developers select

**Table 2. Organizations included in data management analysis, by type.**

Organization type	Percent
Local government	4
State government	17
Federal government	11
International organization	10
Commercial organization	58

process-improvement strategies by determining their current process maturity and identifying the most critical issues to improving their software quality and process.”<sup>10</sup> Similarly, our goal was to aid data management practice improvement by presenting a scale for measuring data management accomplishments. Our assessment results can help data managers identify and implement process improvement strategies by recognizing their data management challenges.

## DATA COLLECTION PROCESS AND RESEARCH TARGETS

Between 2000 and 2006, we assessed the data management practices of 175 organizations. Table 2 provides a breakdown of organization types.

Students from some of our graduate and advanced undergraduate classes largely conducted the assessments. We provided detailed assessment instruction as part of the course work. Assessors used structured telephone and in-person interviews to assess specific organizational data management practices by soliciting evidence of processes, products, and common features. Key concepts sought included the presence of commitments, abilities, measurements, verification, and governance.

Assessors conducted the interviews with the person identified as having the best, firsthand knowledge of organizational data management practices. Tracking down these individuals required much legwork; identifying these individuals was often more difficult than securing the interview commitment.

The assessors attempted to locate evidence in the organization indicating the existence of key process areas within specific data management practices. During the evaluation, assessors observed strict confidentiality—they reported only compiled results, with no mention of specific organizations, individuals, groups, programs, or projects. Assessors and participants kept all information to themselves and observed proprietary rights, including several nondisclosure agreements.

All organizations implement their data management practice in ways that can be classified as one of five maturity model levels, detailed in Table 3 on the next page. Specific evidence, organized by maturity level, helped identify the level of data management practiced.

**Table 3. Data management practice assessment levels.**

Level	Name	Practice	Quality and results predictability
1	Initial	The organization lacks the necessary processes for sustaining data management practices. Data management is characterized as ad hoc or chaotic.	The organization depends entirely on individuals, with little or no corporate visibility into cost or performance, or even awareness of data management practices. There is variable quality, low results predictability, and little to no repeatability.
2	Repeatable	The organization might know where data management expertise exists internally and has some ability to duplicate good practices and successes.	The organization exhibits variable quality with some predictability. The best individuals are assigned to critical projects to reduce risk and improve results.
3	Defined	The organization uses a set of defined processes, which are published for recommended use.	Good quality results within expected tolerances most of the time. The poorest individual performers improve toward the best performers, and the best performers achieve more leverage.
4	Managed	The organization statistically forecasts and directs data management, based on defined processes, selected cost, schedule, and customer satisfaction levels. The use of defined data management processes within the organization is required and monitored.	Reliability and predictability of results, such as the ability to determine progress or six sigma versus three sigma measurability, is significantly improved.
5	Optimizing	The organization analyzes existing data management processes to determine whether they can be improved, makes changes in a controlled fashion, and reduces operating costs by improving current process performance or by introducing innovative services to maintain their competitive edge.	The organization achieves high levels of results certainty.

For each data management process, the assessment used between four and six objective criteria to probe for evidence. Assessed outside the data collection process, the presence or absence of this evidence indicated organizational performance at a corresponding maturity level.

### ASSESSMENT RESULTS

The assessment results reported for the various practice areas show that overall scores are repeatable (level 2) in all data management practice areas.

Figure 3 shows assessment averages of the individual response scores. We used a composite chart to group the averages by practice area. Such groupings facilitate numerous comparisons, which organizations can use to plan improvements to their data management practices.

We present sample results (blue) for an assessed organization (disguised as “Mystery Airline”), whose management was interested in not only how the organization scored but also how it compared to other assessed airlines (red) and other organizations (white).

We grouped 19 individual responses according to the five data management maturity levels in the horizontal bar charts. Most numbers are averages. That is, for an individual organization, we surveyed multiple data management operations, combined the individual assessment results, and presented them as averages. We reported assessments of organizations with only one data management function as integers.

For example, the data program coordination practice area results include:

- Mystery Airline achieved level 1 on responses 1, 2, and 5, and level 2 on responses 3 and 4.
- The airline industry performed above both Mystery Airline and all respondents on responses 1 through 3.
- The airline industry performed below both Mystery Airline and all respondents on response 4, and Mystery Airline performed well below all respondents and just those in the airline industry on response 5.

Figure 3f illustrates the range of results for all organizations surveyed for each data management process—for example, the assessment results for data program coordination ranged from 2.06 to 3.31.

The maturity measurement framework dictates that a data program can achieve no greater rating than the lowest rating achieved—hence the translation to the scores for Mystery Airline of 1, 2, 2, 2, and 2 combining for an overall rating of 1. This is congruent with CMMI application.

Although this might seem a tough standard, the rating reflects the adage that a chain is only as strong as its weakest link. Mature data management programs can’t rely on immature or ad hoc processes in related areas. The lowest rating received becomes the highest possible

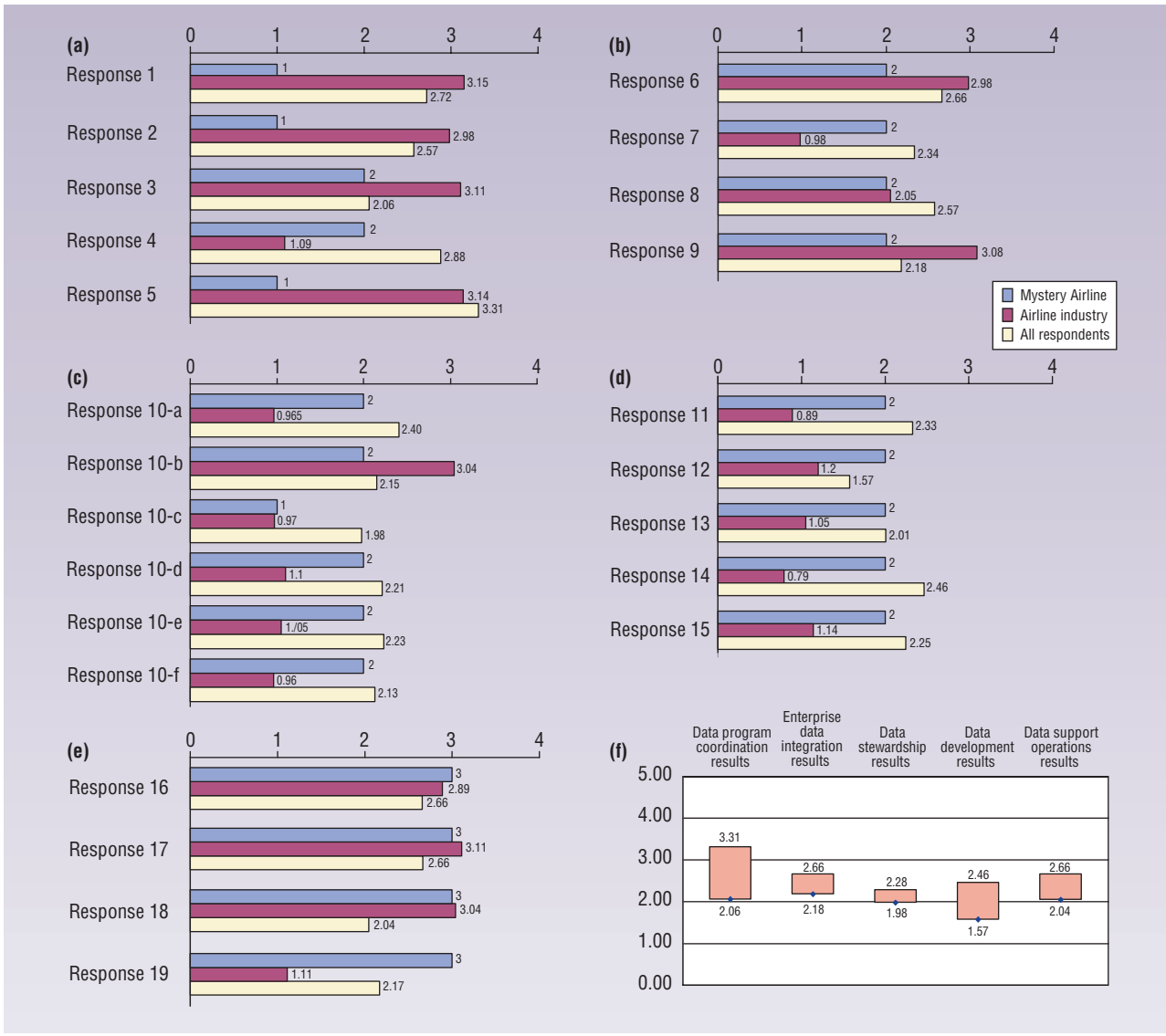


Figure 3. Assessment results useful to Mystery Airline: (a) data program coordination, (b) enterprise data integration, (c) data stewardship, (d) data development, (e) data support organizations, and (f) assessments range.

overall rating. This also explains why many organizations are at level 1 with regard to their software development practices. While the CMMI process results in a single overall rating for the organization, data management requires a more fine-grained feedback mechanism. Knowing that some data management processes perform better than others can help an organization develop incentives as well as a roadmap for improving individual ratings.

Taken as a whole, these numbers show that no data management process or subprocess measured on average higher than the data program coordination process, at 3.31. It's also the only data management process that performed on average at a defined level (greater than 3). The results show a community that is approaching the ability to repeat its processes across all of data management.

### Results analysis

Perhaps the most important general fact represented in Figure 3 is that organizations gave themselves relatively low scores. The assessment results are based on self-reporting and, although our 15-percent validation sample is adequate to verify accurate industry-wide assessment results, 85 percent of the assessment is based on facts that were described but not observed. Although direct observables for all survey respondents would have provided valuable confirming evidence, the cost of such a survey and the required organizational access would have been prohibitive.

We held in-person, follow-up assessment validation sessions with about 15 percent of the assessed organizations. These sessions helped us validate the collection method and refine the technique. They also let us gauge the assessments' accuracy.

**Table 4. Assessment scores adjusted for self-reporting inflation.**

Response	Adjusted average
1	1.72388
2	1.57463
3	1.0597
4	1.8806
5	2.31343
6	1.66418
7	1.33582
8	1.57463
9	1.1791
10 a	1.40299
10 b	1.14925
10 c	0.97761
10 d	1.20896
10 e	1.23134
10 f	1.12687
11	1.32836
12	0.57463
13	1.00746
14	1.46269
15	1.24627
16	1.65672
17	1.66418
18	1.04478
19	1.17164

Although the assessors strove to accurately measure each subprocess's maturity level, some interviews inevitably were skewed toward the positive end of the scale. This occurred most often because interviewees reported on milestones that they wanted to or would soon achieve as opposed to what they had achieved. We suspected, and confirmed during the validation sessions, that responses were typically exaggerated by one point on the five-point scale.

When we factor in the one-point inflation, the numbers in Table 4 become important. Knowing that the bar is so low will hopefully inspire some organizations to invest in data management. Doing so might give them a strategic advantage if the competition is unlikely to be making a similar investment.

The relatively low scores reinforce the need for this data management assessment. Based on the overall scores in the data management practice areas, the community receives five Ds. These areas provide immediate targets for future data management investment.

### WHERE ARE WE NOW?

We address our original research objectives according to our two goal categories.

### Community descriptive research goals

First, we wanted to determine the range of practices within the data management community. A wide range of such practices exists. Some organizations are strong in some data management practices and weak in others (the range of practice is consistently inconsistent). The wide divergence of practices both within and between organizations can dilute results from otherwise strong data management programs. The assessment's applicability to longitudinal studies remains to be seen; this is an area for follow-up research. Although researchers might undertake formal studies of such trends in the future, evidence from ongoing assessments suggests that results are converging. Consequently, we feel that our sample constitutes a representation of community-wide data management practices.

Next, we wanted to know whether the distribution of practices informs us specifically about the various stages of organizational data management maturity. The assessment results confirm the framework's utility, as do the postassessment validation sessions. Building on the framework, we were able to specify target characteristics and objective measurements. We now have better information as to what comprises the various stages of organizational data management practice maturity. Organizations do clump together into the various maturity stages that Nolan originally described. We can now determine the investments required to predictably move organizations from one data management maturity level to another.

Finally, we wanted to determine in what areas the community data management practices are weak, average, and strong. Figure 4 shows an average of unadjusted rates summarizing the assessment results. As the figure shows, the data management community reports itself relatively and perhaps surprisingly strong in all five major data management processes when compared to the industry averages for software development. The range and averages indicate that the data management community has more mature data program coordination processes, followed by organizational data integration, support operations, stewardship, and then data development. The relatively lower data development scores might suggest data program coordination implementation difficulties.

### Self-improvement research goals

Our first objective was to produce results that would help the community better understand current best practices. Organizations can use the assessment results to compare their specific performance against others in their industry and against the community results as a whole. Quantities and groupings indicate the relative state and robustness of the best practices within each process. Future research can use this information to identify specific practices that can be shared with the



community. Further study of these areas will provide leverageable benefits.

Next, we wanted to determine how the assessment informs our standing as a technical community. Our research gives some indication of the claimed current state of data management practices. However, given the validation session results, we believe that it's best to caution readers that the numbers presented probably more accurately describe the intended state of the data management community.

As it turns out, the relative number of organizations above level 1 for both software and data management are approximately the same, but a more detailed analysis would be helpful. Given the belief that investment in software development practices will result in significant improvements, it's appropriate to anticipate similar benefits from investments in data management practices.

Finally, we hoped to gain information useful for developing a roadmap for improving current practice. Organizations can use the survey assessment information to develop roadmaps to improve their individual data management practices. Mystery Airline, for example, could develop a roadmap for achieving data management improvement by focusing on enterprise data integration, data stewardship, and data development practices.

### SUGGESTIONS FOR FUTURE RESEARCH

Additional research must include a look at relationships between data management practice areas, which could indicate an efficient path to higher maturity levels. Research should also explore the success or failure of previous attempts to raise the maturity levels of organizational data management practices.

One of our goals was to determine why so many organizational data management practices are below expectations. Several current theses could spur investigation of the root causes of poor data management practices. For example,

- Are poor data management practices a result of the organization's lack of understanding?
- Does data management have a poor reputation or track record in the organization?
- Are the executive sponsors capable of understanding the subject?
- How have personnel and project changes affected the organization efforts?

Our assessment results suggest a need for a more formalized feedback loop that organizations can use to

	Initial	Repeatable			Defined
Data program coordination		2.06	2.71	3.31	
Enterprise data integration		2.18	2.44	2.66	
Data stewardship		1.98	2.18	2.40	
Data development	1.57	2.12	2.46		
Data support operations		2.04	2.38	2.66	

Figure 4. Average of unadjusted rates for the assessment results, by process.

improve their data management practices. Organizations can use this data as a baseline from which to look for, describe, and measure improvements in the state of the practice. Such information can enhance their understanding of the relative development of organizational data management. Other investigations should probe further to see if patterns exist for specific industry or business focus types.

Building an effective business case for achieving a certain level of data management is now easier. The failure to adequately address enterprise-level data needs has hobbled past efforts.<sup>4</sup> Data management has, at best, a business-area focus rather than an enterprise outlook. Likewise, applications development focuses almost exclusively on line-of-business needs, with little attention to cross-business-line data integration or enterprise-wide planning, analysis, and decision needs (other than within personnel, finance, and facilities management). In addition, data management staff is inexperienced in modern data management needs, focusing on data management rather than metadata management and on syntaxes instead of semantics and data usage.

Few organizations manage data as an asset. Instead, most consider data management a maintenance cost. A small shift in perception (from viewing data as a cost to regarding it as an asset) can dramatically change how an organization manages data. Properly managed data is an organizational asset that can't be exhausted. Although data can be polluted, retired, destroyed, or become obsolete, it's the one organizational resource that can be repeatedly reused without deterioration, provided that the appropriate safeguards are in place. Further, all organizational activities depend on data.

To illustrate the potential payoff of the work presented here, consider what 300 software professionals applying software process improvement over an 18-year period achieved:<sup>16</sup>

- They predicted costs within 10 percent.
- They missed only one deadline in 15 years.
- The relative cost to fix a defect is 1X during inspection, 13X during system testing, and 92X during operation.

- Early error detection rose from 45 to 95 percent between 1982 and 1993.
- Product error rate (measured as defects per 1,000 lines of code) dropped from 2.0 to 0.01 between 1982 and 1993.

If improvements in data management can produce similar results, organizations should increase their maturity efforts. ■

---

### Acknowledgments

We thank Graham Blevins, David Rafner, and Santa Susarapu for their assistance in preparing some of the reported data. We are greatly indebted to many of Peter Aiken's classes in data reengineering and related topics at Virginia Commonwealth University for the careful work and excellent results obtained as a result of their various contributions to this research. This article also benefited from the suggestions of several anonymous reviewers. We also acknowledge the helpful, continuing work of Brett Chaplin at Allstate in collecting, applying, and assessing CMMI-related efforts.

---

### References

1. M. Blaha, "A Retrospective on Industrial Database Reverse Engineering Projects—Parts 1 & 2," *Proc. 8th Working Conf. Reverse Eng.*, IEEE Press, 2001, pp. 147-164.
2. J. Zachman, "A Framework for Information Systems Architecture," *IBM Systems J.*, vol. 26, 1987, pp. 276-292.
3. P.H. Aiken, "Keynote Address to the 2002 DAMA International Conference: Trends in Metadata," *Proc. 2002 DAMA Int'l/Metadata Conf.*, CD-ROM, Wilshire Conf., 2002, pp. 1-32.
4. B. Parker, "Enterprise Data Management Process Maturity," *Handbook of Data Management*, S. Purba, ed., Auerbach Publications, CRC Press, 1999, pp. 824-843.
5. M. Gillenson, "The State of Practice of Data Administration—1981," *Comm. ACM*, vol. 25, no. 10, 1982, pp. 699-706.
6. M. Gillenson, "Trends in Data Administration," *MIS Quarterly*, Dec. 1985, pp. 317-325.
7. M. Gillenson, "Database Administration at the Crossroads: The Era of End-User-Oriented, Decentralized Data Processing," *J. Database Administration*, Fall 1991, pp. 1-11.
8. R.L. Keeney, *Value-Focused Thinking—A Path to Creative Decisionmaking*, Harvard Univ. Press, 1992.
9. R. Nolan, "Managing the Crisis in Data Processing," *Harvard Business Rev.*, Mar./Apr. 1979, pp. 115-126.
10. Carnegie Mellon Univ. Software Eng. Inst., *Capability Maturity Model: Guidelines for Improving the Software Process*, 1st ed., Addison-Wesley Professional, 1995.
11. M.C. Paulk and B. Curtis, "Capability Maturity Model, Version 1.1," *IEEE Software*, vol. 10, 1993, pp. 18-28.
12. D.R. Goldenson and D.L. Gibson, "Demonstrating the Impact and Benefits of CMM: An Update and Preliminary Results," special report CMU/SEI-2003-SR-009, Carnegie Mellon Univ. Software Eng. Inst., 2003, pp. 1-55.
13. C. Billings and J. Clifton, "Journey to a Mature Software Process," *IBM Systems J.*, vol. 33, 1994, pp. 46-62.
14. C.C. Hauer, "Data Management and the CMM/CMMI: Translating Capability Maturity Models to Organizational Functions," presented at National Defense Industrial Assoc. Technical Information Division Symp., 2003; [www.dtic.mil/ndia/2003technical/hauer1.ppt](http://www.dtic.mil/ndia/2003technical/hauer1.ppt).
15. W. Schnider and K. Schwinn, "Der Reifegrad des Datenmanagements" [The Data Management Maturity Model], KPP Consulting; [www.kpp-consulting.ch/downloadbereich/DM%20Maturity%20Model.pdf](http://www.kpp-consulting.ch/downloadbereich/DM%20Maturity%20Model.pdf), 2004 (in German).
16. H. Krasner, J. Pyles, and H. Wohlwend, "A Case History of the Space Shuttle Onboard Systems Project," Technology Transfer 94092551A-TR, Sematech, 31 Oct. 1994.

*Peter Aiken is an associate professor of information systems at Virginia Commonwealth University and founding director of Data Blueprint. His research interests include data and systems reengineering. Aiken received a PhD in information technology from George Mason University. He is a senior member of the IEEE, the ACM, and the Data Management Association (DAMA) International. Contact him at [peter@datablueprint.com](mailto:peter@datablueprint.com).*

*M. David Allen is chief operating officer of Data Blueprint. His research interests include data and systems reengineering. Allen received an MS in information systems from Virginia Commonwealth University. He is a member of DAMA. Contact him at [mda@datablueprint.com](mailto:mda@datablueprint.com).*

*Burt Parker is an independent consultant based in Washington, DC. His technical interests include enterprise data management program development. Parker received an MBA in operations research/systems analysis (general systems theory) from the University of Michigan. He is a member of DAMA. Contact him at [parkerbg@comcast.net](mailto:parkerbg@comcast.net).*

*Angela Mattia is a professor of information systems at J. Sergeant Reynolds Community College. Her research interests include data and systems reengineering and maturity models. Mattia received an MS in information systems from Virginia Commonwealth University. She is a member of DAMA. Contact her at [amattia@jsr.vccs.edu](mailto:amattia@jsr.vccs.edu).*