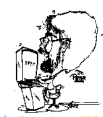
"Practice isn't the thing you do once you're good. It's the thing you do that makes you good." - Malcolm Gladwell

Data Strategy Best Practices



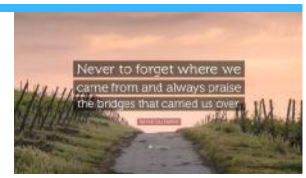
peter.aiken@anythingawesome.com +1.804.382.5957

Peter Aiken, Ph.D.

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

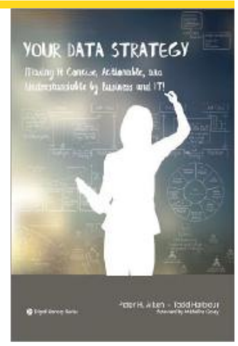
$1,500,000,000.00 USD

Context



- Strategy
  - Inherently a repetitive process that can be easily improved
- Dependency
  - Data strategy exists to support organizational strategy
- Evolution
  - Focus on improving data capabilities
- Output
  - Plans are of limited value anyway and always discount obstacles
- Overemphasizing Technology
  - People and process challenges are 95% of the problem
- Nirvana
  - Q: How do I get to Carnegie Hall?
  - A: Practice Practice Practice

Program Overview

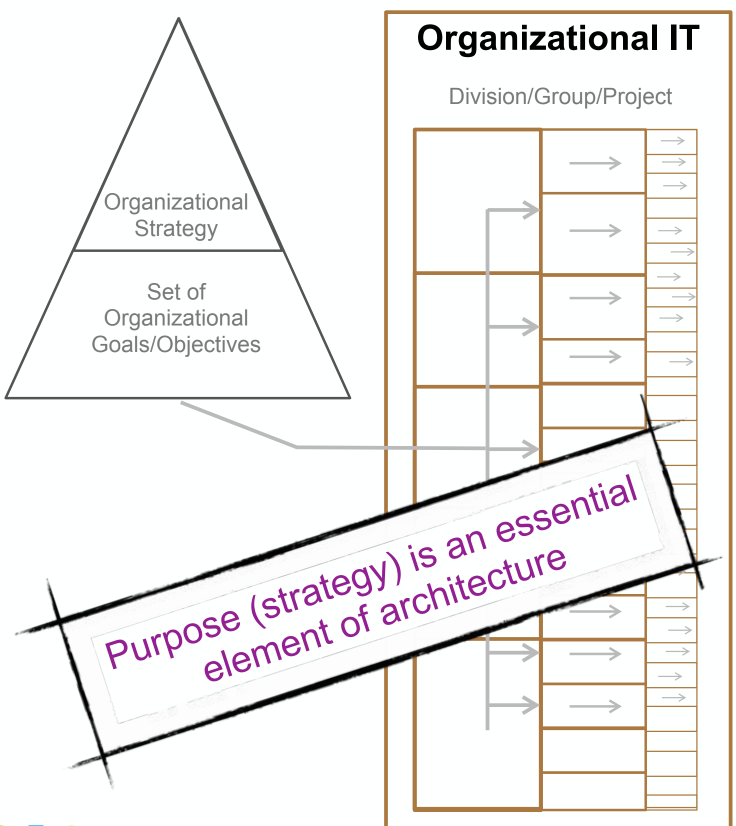


- A data strategy specifies how data assets are to be used to support the organizational strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- A data strategy is necessary for effective data governance
  - Improve your organization's data
  - Improve the way people use their data
  - Improving how people use data to support their organizational strategy
- Effective Data Strategy Prerequisites
  - Lack of organizational readiness
  - Failure to compensate for the lack of data competencies
  - Eliminating the barriers to leveraging data, the seven deadly data sins
- Data Strategy Development Phase II–Iterations
  - Lather, rinse, repeat
  - A balanced approach is required
  - Establish various data value chains
- Q&A

Data Strategy Best Practices

https://ctb.ku.edu/en/table-of-contents/structure/strategic-planning/develop-strategies/main

Strategy is Difficult to Perceive at the IT Project Level

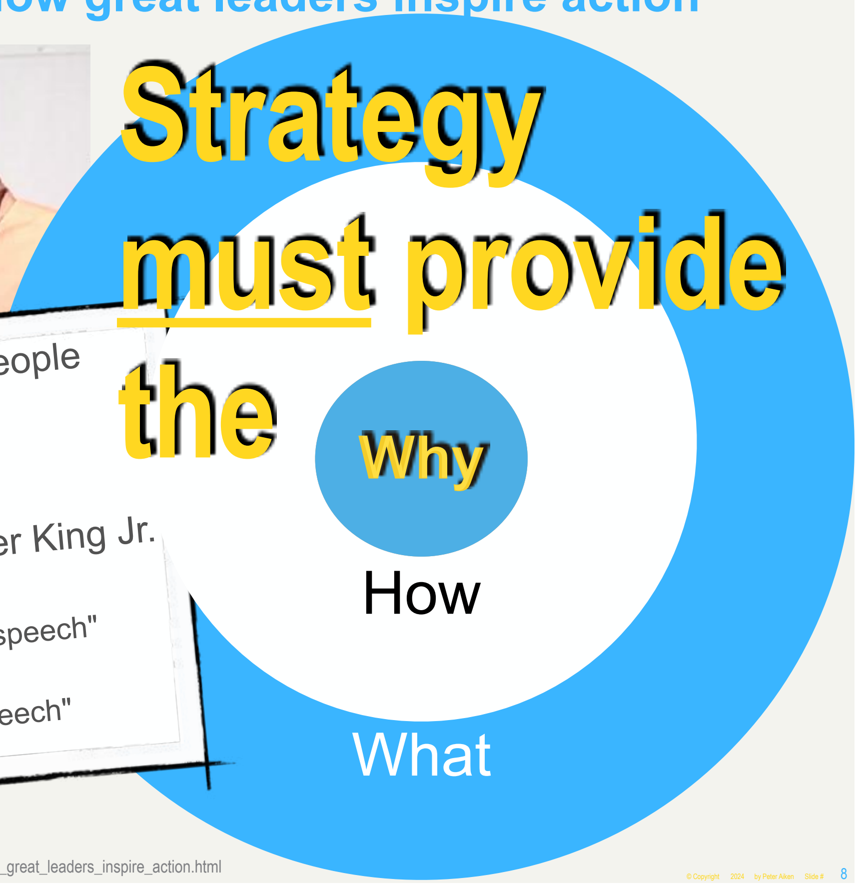


- If they exist ...
- A singular organizational strategy and set of goals/ objectives ...
- Are not perceived as such at the project level and ...
- What does exist is confused, inaccurate, and incomplete
- IT projects do not well reflect organizational strategy

https://www.cio.com/article/3088208/leadership-management/how-to-avoid-a-digital-strategy-failure.html

Simon Sinek: How great leaders inspire action

- What motivates people
  - is not what you do,
  - It is **why** you do it (for example)
- Rev. Martin Luther King Jr. gave the
  - "I have a **dream** speech" (not the)
  - "I have a plan speech"



http://www.ted.com/talks/simon_sinek_how_great_leaders_inspire_action.html

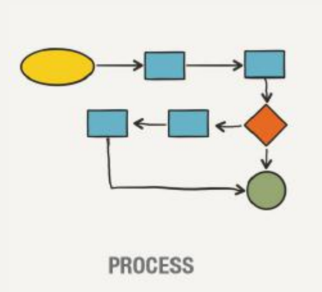
What is Strategy?

strat·e·gy
/ˈstratəjē/

noun

1. a plan of action or policy designed to achieve a major or overall aim.
   "time to develop a coherent economic strategy"
   synonyms: master plan, grand design, game plan, plan (of action), action plan, policy, program; More

- Current use derived from military
  - **a pattern in a stream of decisions**
    [Henry Mintzberg]



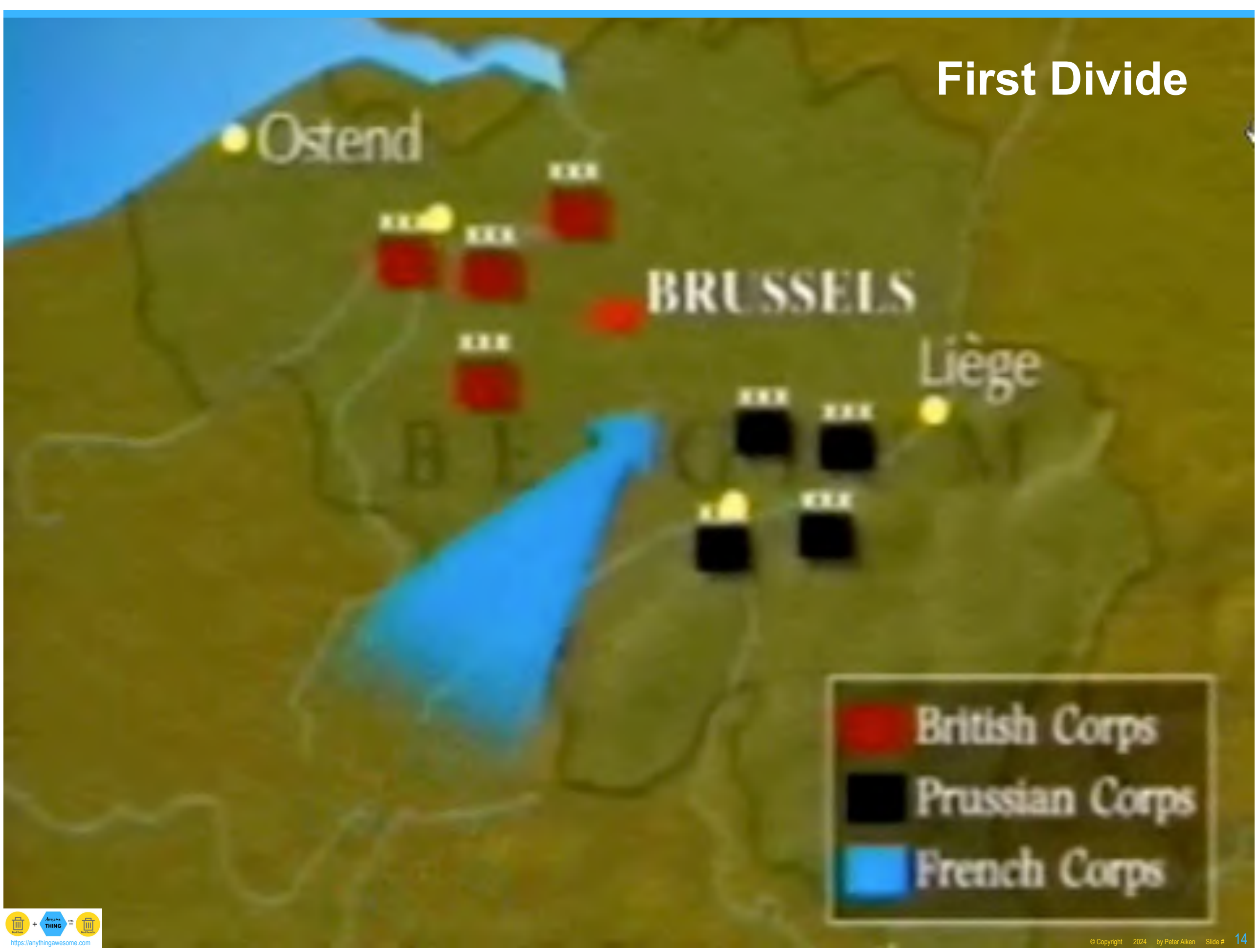
Former Walmart Business Strategy

Every Day Low Price

Strategy in Action: Napoleon faces a larger enemy

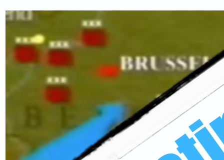
- Question?
  - How do I defeat the competition when their forces are bigger than mine?
- Answer:
  - Divide and conquer!
  - “a pattern in a stream of decisions”


LINES OF SUPPLY
Supply Line Metadata
(as part of a divide and conquer strategy)
Ostend
BRUSSELS
Liège
British Corps
Prussian Corps
French Corps
https://anythingawesome.com
© Copyright 2024 by Peter Aiken Slide # 13


First Divide
Ostend
BRUSSELS
Liège
British Corps
Prussian Corps
French Corps
https://anythingawesome.com
© Copyright 2024 by Peter Aiken Slide # 14

Complex Strategy

- First
  - Hit both armies hard at just the right spot



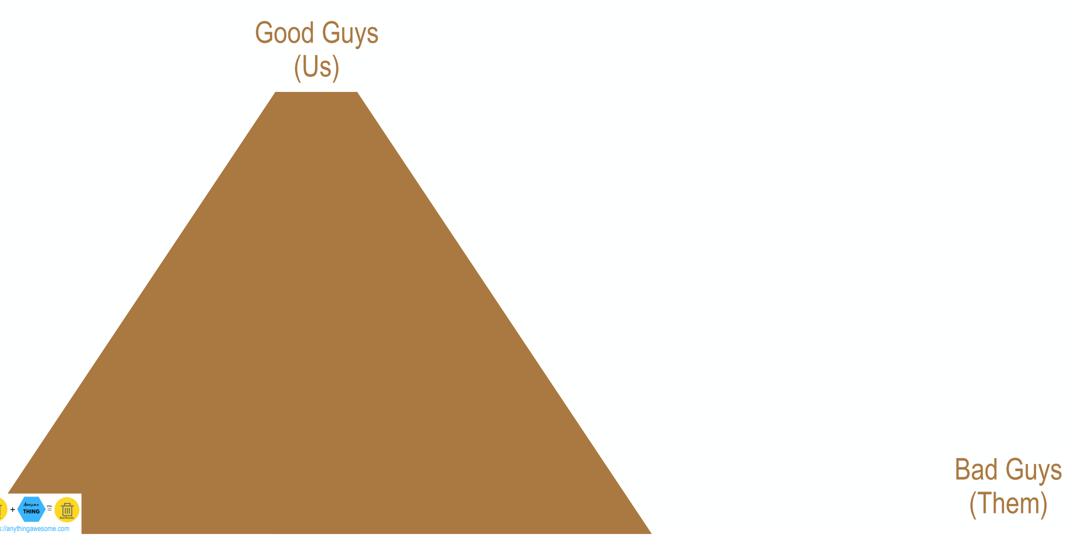
- Then
  - Turn right defeat

- –urn left and defeat the British

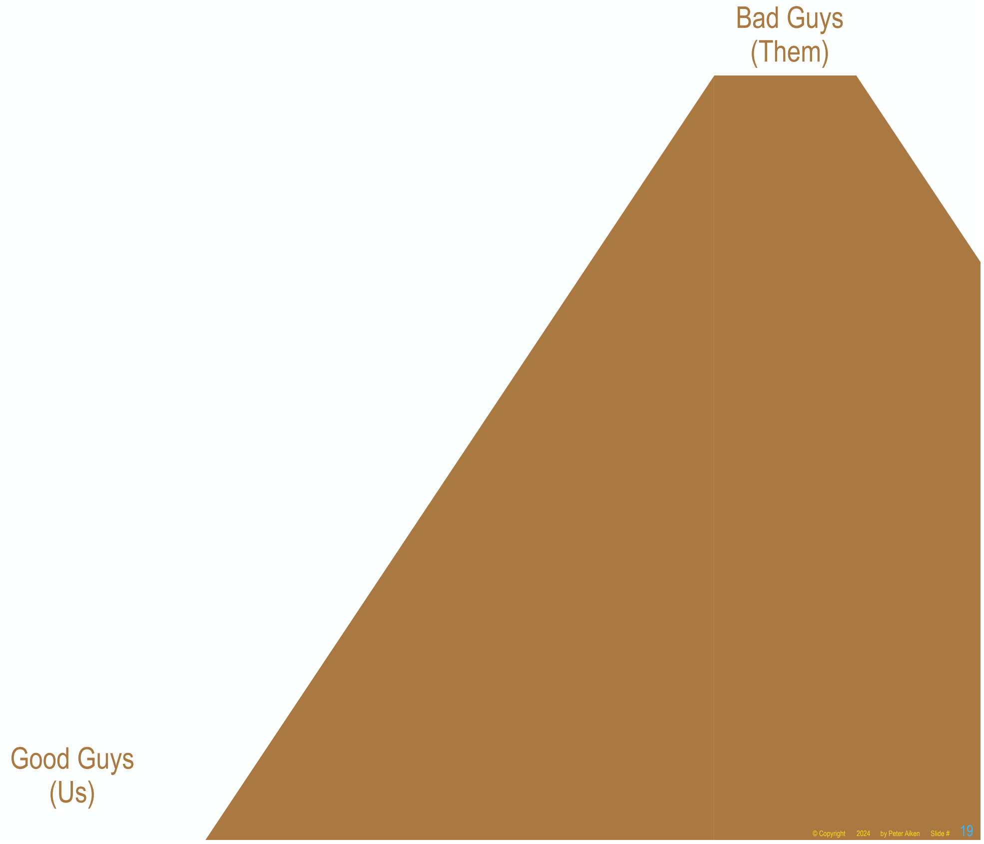
While someone is shooting at you!

Contextually Important Strategy Example 1

Contextually Important Strategy Example 2



Contextually Important Strategy Example 3



Strategy Guides Workgroup Activities

A pattern in a stream of decisions

Strategy that winds up only on a shelf is not useful

General Dwight D. Eisenhower

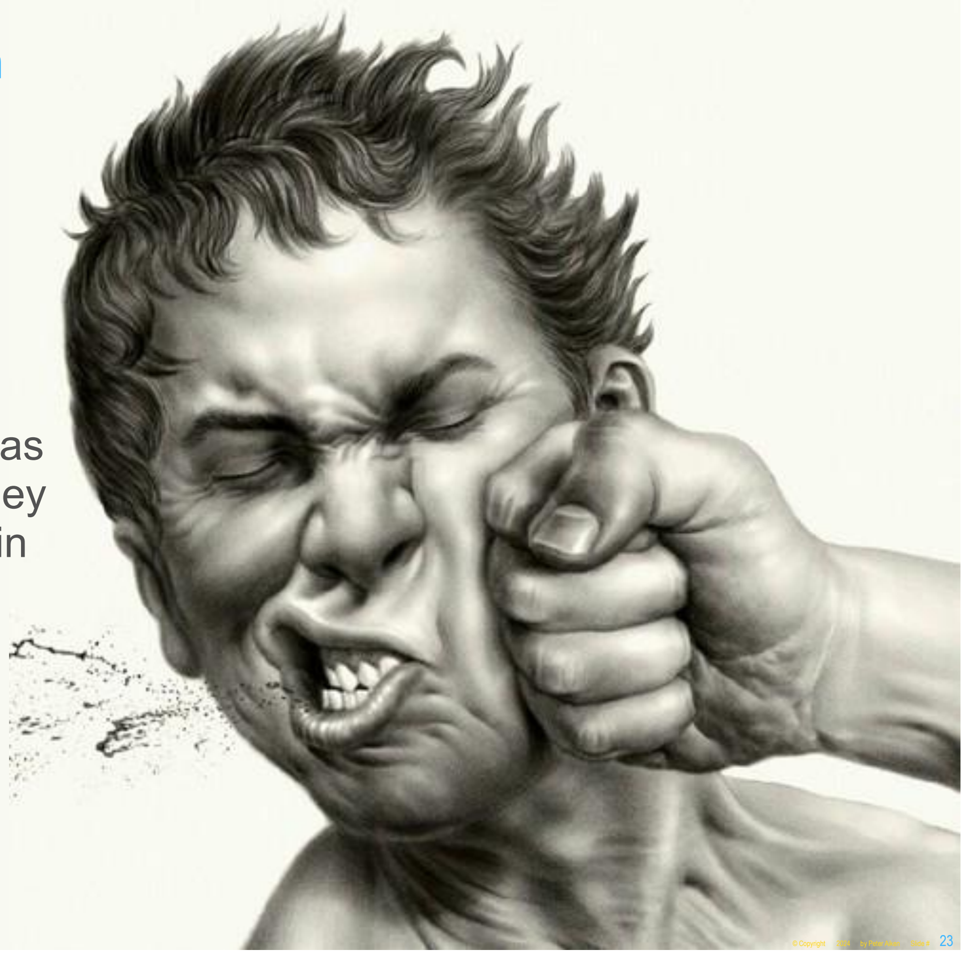
Military Plan

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/

Mike Tyson



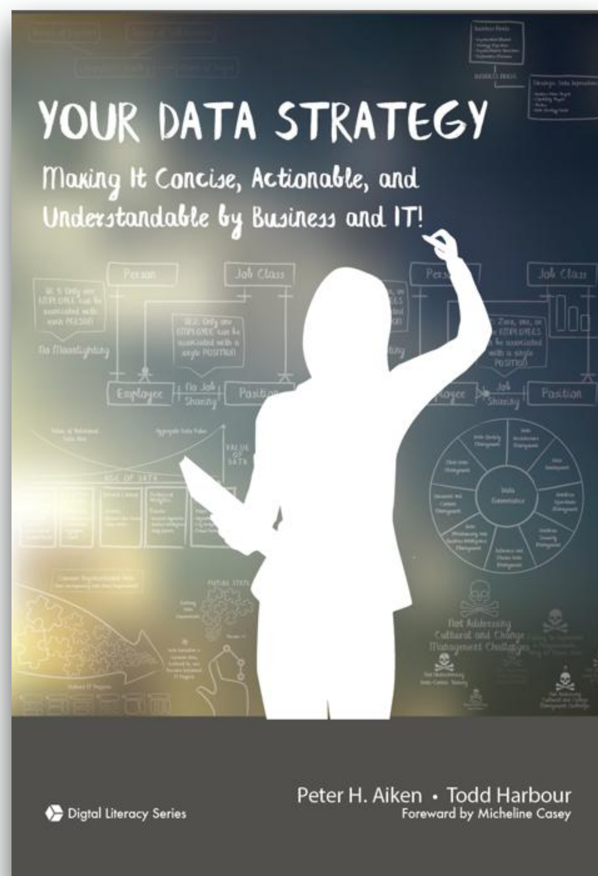
"Everybody has a plan until they get punched in the face."

http://f--f.info/?p=23071

Your Data Strategy

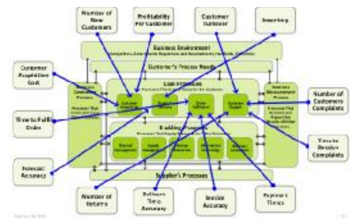
- Highest level data guidance available ...
- Focusing data activities on business-goal achievement ...
- Providing guidance when faced with a stream of decisions or uncertainties

> - Data strategy most usefully articulates how data can be best used to support organizational strategy
> - This usually involves a balance of remediation and proactive measures



Data Strategy Measures

- Effectiveness
  - Over time



- Volume (length)
  - Should be not a whole lot longer than the organizational strategy
    https://www.gartner.com/en/webinars/3994588/the-art-of-the-1-page-strategy-storytelling-enables-business-gro
- Versions
  - Should be sequential (with score keeping)

- Understanding
  - Common agreement can be measured

Information Management Strategy On A Page

Statement of Information Management Strategy: Shift the focus of IT investment and skills toward information management with the goal of providing employees with attainable and useful information and boosting their capability to exploit that information for competitive advantage.

State of IM in 20XX

Top 5-7 Metrics Describing the Initial State

- Percentage of IT budget devoted to information and analytics projects = 23%
- Percentage of budget spent on employee capability < 5%
- Percentage of information subjects targeted for harmonization and integration > 80%
- Target number of analytic tools = 1-3
- Percentage of employees who are informed skeptics = 38%

Top 5-7 IM Initiatives

1. Launch an information management and analytics center of excellence.
2. Redesign IT's opportunity-identification process to make it proactive and informed by observation of distinct employee segments.
3. Identify analytic capabilities used by employees and offer a portfolio of tools to meet those needs.
4. Develop and hire usability and inte... design skills in IT.
5. Coach employees to boost their a... and foster informed skepticism.
6. Harmonize and integrate a sma... subjects where there is greates...

Top 5-7 Underlying Beliefs and

1. The number of opportunities to drive grow... information management will equal or outstrip the opportunities for process automation.
2. Many of our employees lack the skills and judgment to use information effectively for decision making.
3. Not all information needs to be harmonized or integrated at enterprise level. Similarly, some information needs higher levels of quality than others.
4. Our business partners will take the lead in information stewardship.
5. Employee reliance on external information sources and on unstructured information will continue to rise.

State of IM in 20YY

Top 5-7 Metrics Describing the End State

- Percentage of IT budget devoted to information and analytics projects = 40-50%
- ...ntage of budget spent ...ability > 10%
- ...tion
- Target number of analytic tools = 8-12
- Percentage of employees who are informed skeptics = 59%

Aspirational

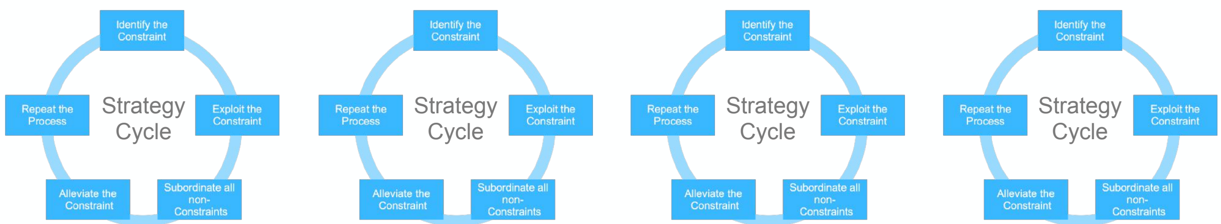
Example courtesy of Dr. Chris Bradley - chris.bradley@dmadvisors.co.uk

Planning Options

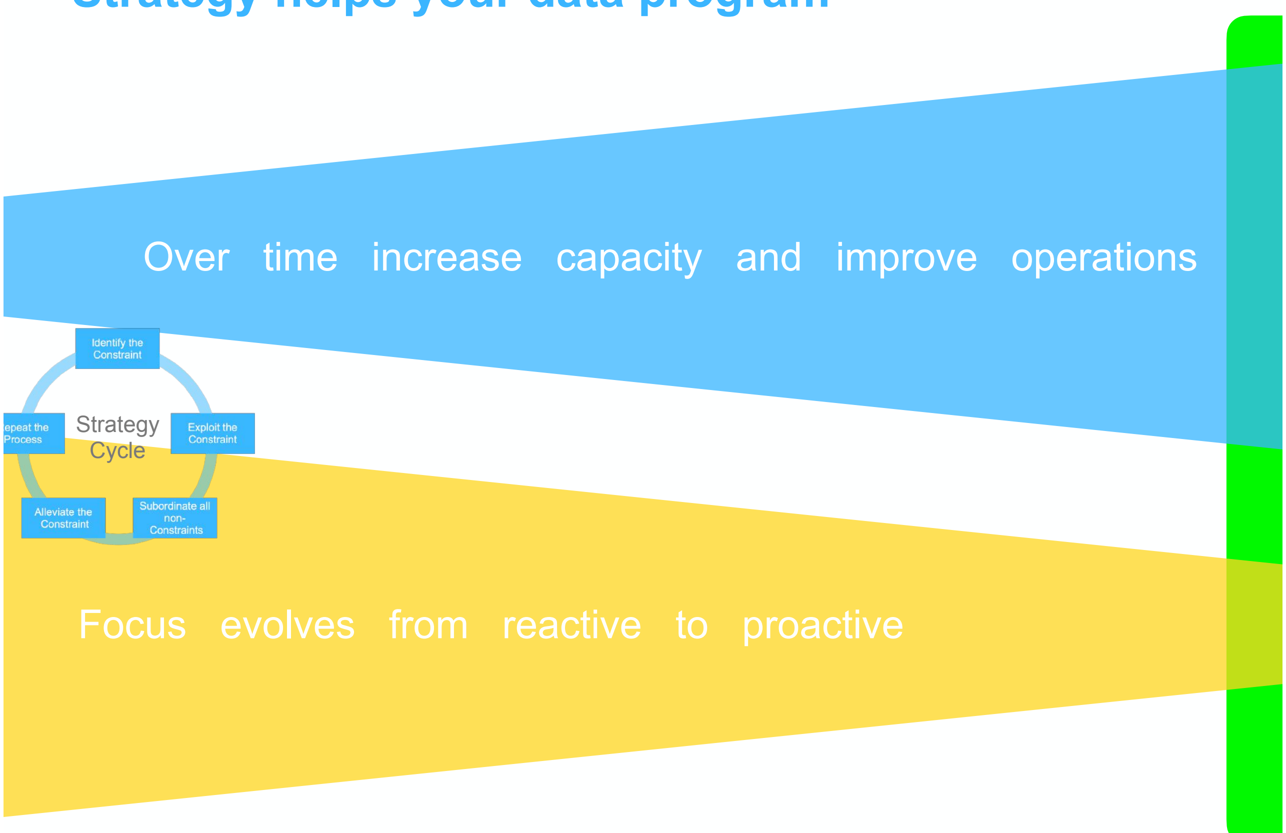
- Plan the entire process before beginning
  - One attempt

Note: Numerous upfront assumptions are required because plans must be detailed, specifying end objectives

- Use iterative strategy cycles
  - Incorporate corrective feedback on initial assumptions



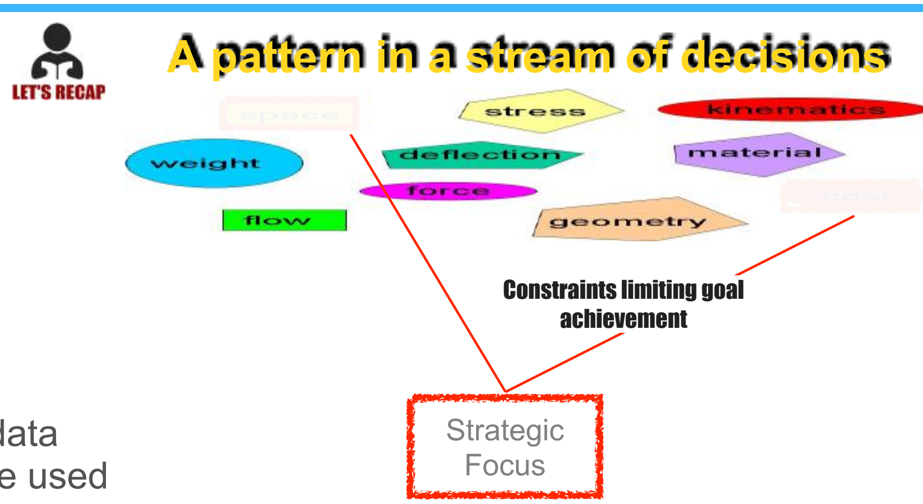
Strategy helps your data program



Other recent data "strategies"

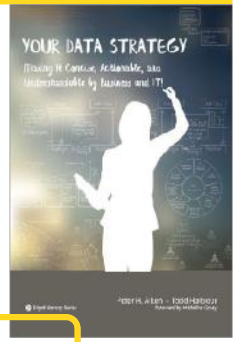
- Big Data
- Data Science
- Analytics
- SAP
- Microsoft
- Google
- AWS
- ...

Recap



- A data strategy specifies how data assets are to be used to support strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- Strategy evolves periodically

Program Overview



- A data strategy specifies how data assets are to be used to support the organizational strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- A data strategy is necessary for effective data governance
  - Improve your organization's data
  - Improve the way people use their data
  - Improving how people use data to support their organizational strategy
- Effective Data Strategy Prerequisites
  - Lack of organizational readiness
  - Failure to compensate for the lack of data competencies
  - Eliminating the barriers to leveraging data, the seven deadly data sins
- Data Strategy Development Phase II–Iterations
  - Lather, rinse, repeat
  - A balanced approach is required
  - Establish various data value chains
- Q&A

Data Strategy Best Practices

7 Data Governance Definitions

- The formal orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset – **The MDM Institute**
- A convergence of data quality, data management, business process management, and risk management surrounding the handling of data in an organization – **Wikipedia**
- A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods – **Data Governance Institute**
- The execution and enforcement of authority over the management of data assets and the performance of data functions – **KiK Consulting**
- A quality control discipline for assessing, managing, using, improving, monitoring, maintaining, and protecting organizational information – **IBM Data Governance Council**
- Data governance is the formulation of policy to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions – **Sunil Soares**
- The exercise of authority and control over the management of data assets – **DM BoK**

Elevator Pitch

An **elevator pitch**, **elevator speech**, or **elevator statement** is a short description of an idea, product, or company that explains the concept in a way such that any listener can understand it in a short period of time.
(Wikipedia)

What is Data Governance?

Go Ask Anyone!™

Managing Data with Guidance

Would you want your sole, non-depletable, non-degrading, durable, strategic asset managed <u>without</u> guidance?

What is Data Governance?

Managing Data Decisions with Guidance

Would you want your sole, non-depletable, non-degrading, durable, strategic asset managed <u>without</u> guidance?

Go Ask Anyone!™

https://anythingawesome.com

Sample from: https://artist.com/kathy-linden/on-outside-looking-in/?artid=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

https://anythingawesome.com

Data Assets Win!

Asset: A resource controlled by the organization as a result of past events or transactions and from which future economic benefits are expected to flow [Wikipedia]

- Today, data is the most powerful, yet underutilized and poorly managed organizational asset
- Data is your
  - Sole (only)
  - Non-depletable
  - Non-degrading
  - Durable
  - Strategic
- Asset
  - Data is the new oil!
  - Data is the new (s)oil!
  - Data is the new bacon!
- As such, data deserves:
  - It's own strategy
  - Attention on par with similar organizational assets
  - Professional ministration to make up for past neglect

	Data Assets	Financial Assets	Real Estate Assets	Inventory Assets
Non-depletable	Available for subsequent use	Can be used up		Can be used up
Non-degrading	√	√	Can degrade over time	Can degrade over time
Durable	Non-taxed		√	√
Strategic Asset	√	√	√	√

2020 American Airlines market value ~ $6b
AAdvantage valued between $19.5-$31.5
2020 United market value ~ 9$b
MileagePlus ~ $22b

https://www.forbes.com/sites/advisor/2020/07/15/how-airlines-make-billions-from-monetizing-frequent-flyer-programs/?sh=66da87a614e9

https://anythingawesome.com

https://anythingawesome.com

Separating the Wheat from the Chaff

Is well organized data worth more?

Pre-Information Age Metadata

- Examples of information architecture achievements that happened well before the information age:
  - Page numbering
  - Alphabetical order
  - Table of contents
  - Indexes
  - Lexicons
  - Maps
  - Diagrams

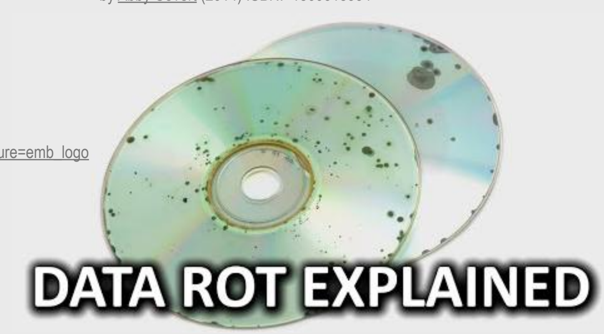
"While we can arrange things with the intent to communicate certain information, we can't actually make information. Our users do that for us."

Example from: *How to make sense of any mess* by Abby Covert (2014) ISBN: 1500615994

https://www.youtube.com/watch?v=60oD1TDzAXQ&feature=emb_logo

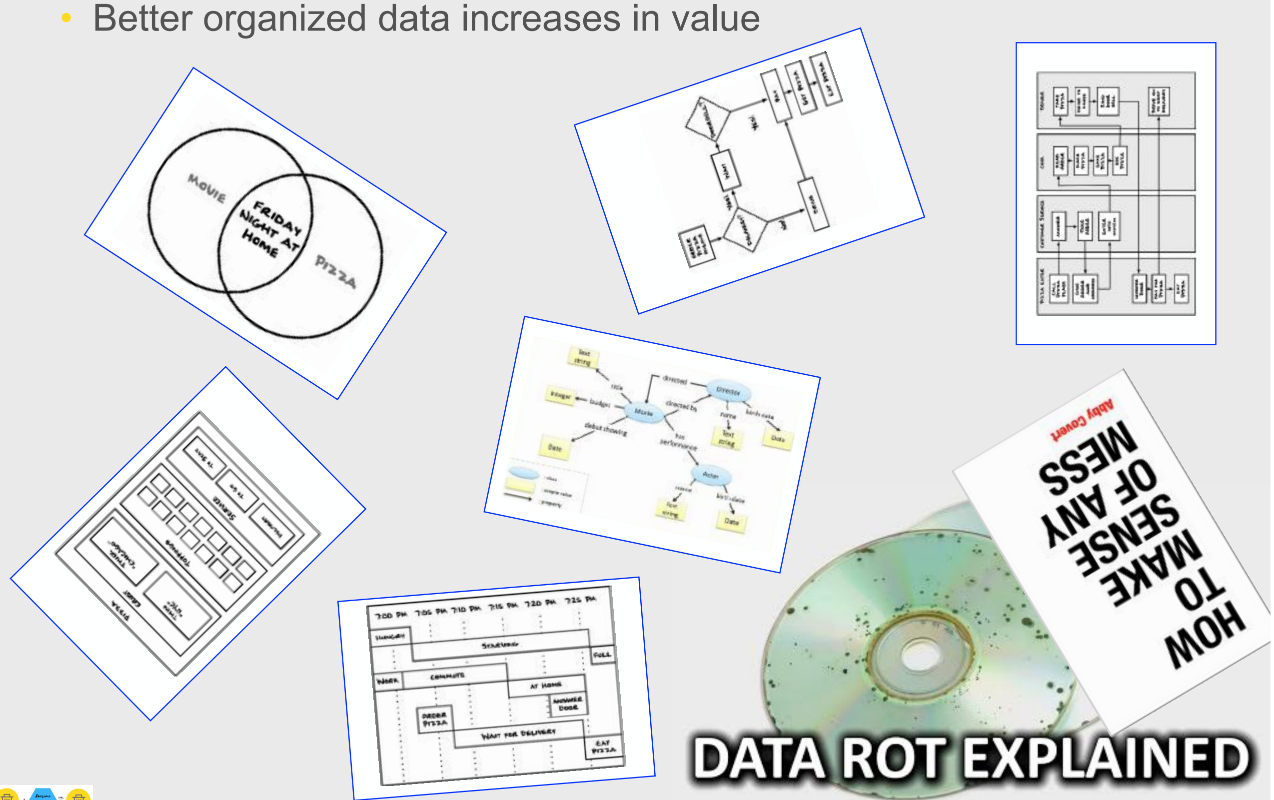
https://www.youtube.com/watch?v=r10Sod44rME&t=1s

https://www.youtube.com/watch?v=XD2OkDPAl6s

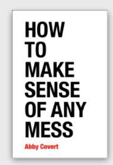


Remove the structure and things fall apart rapidly

- Better organized data increases in value

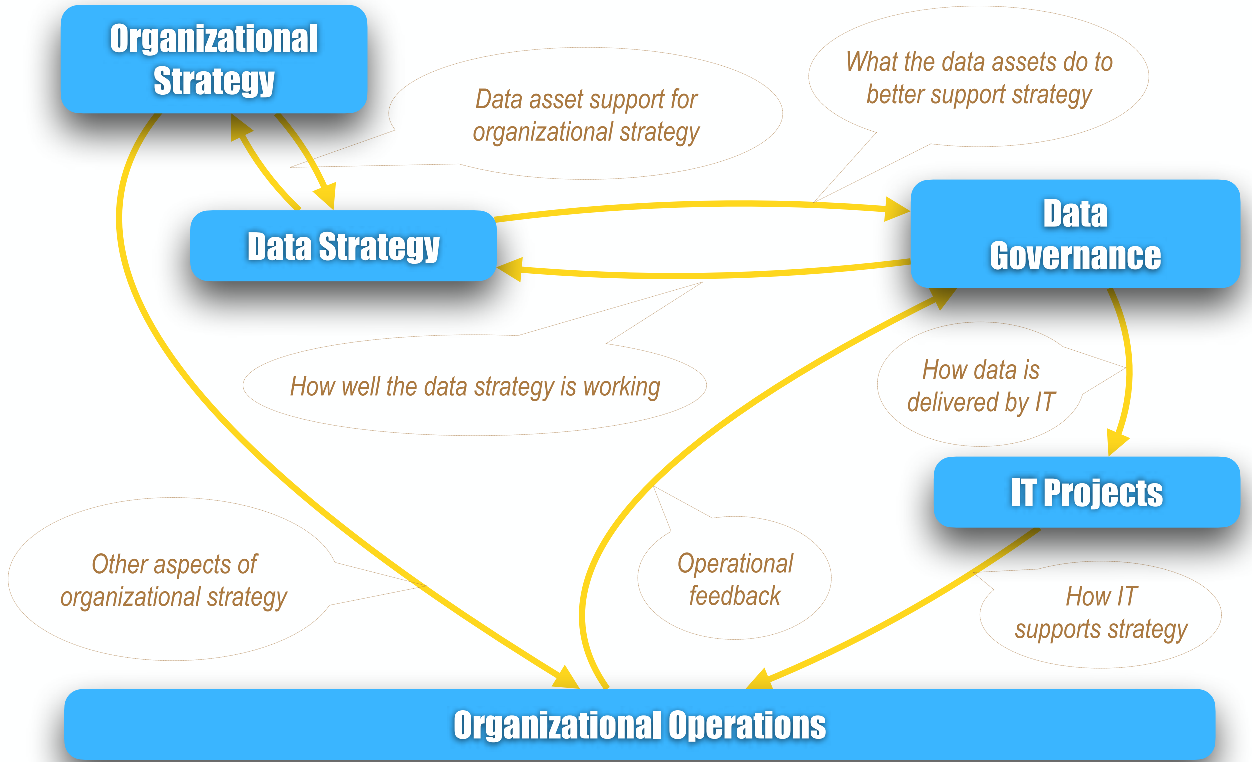


Separating the Wheat from the Chaff

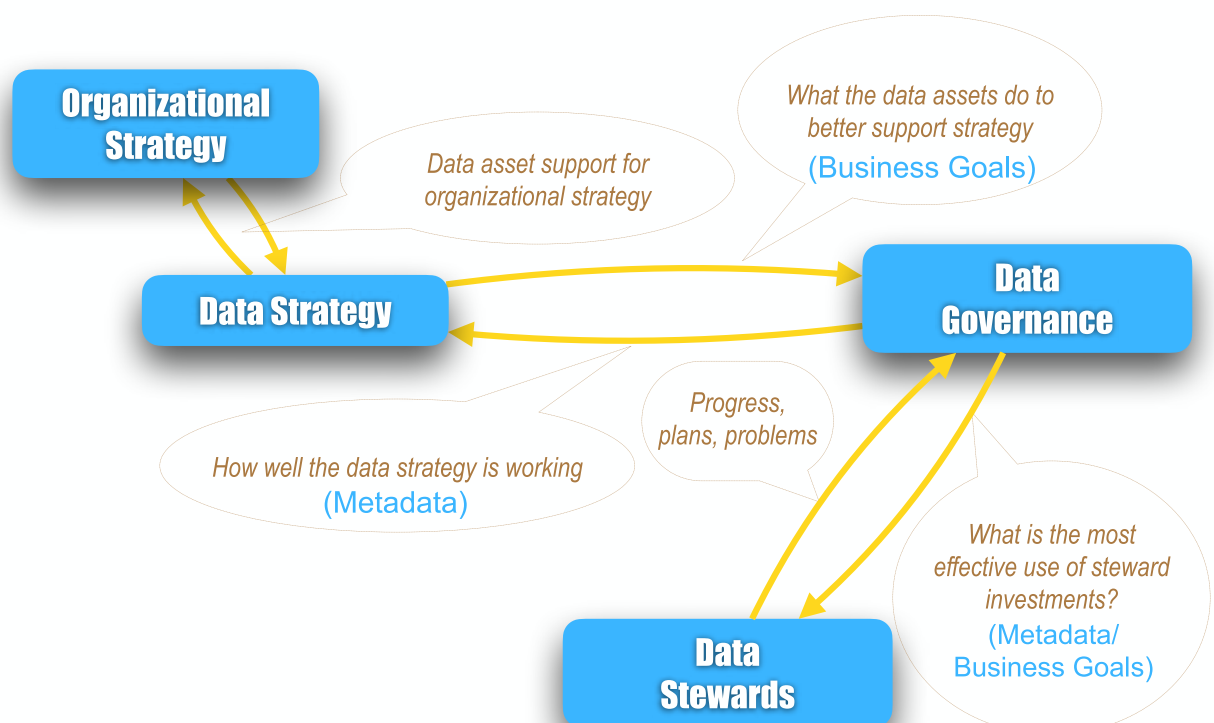


- Better organized data increases in value
- Poor data management practices are costing organizations money/time/effort
- 80% of organizational data is **ROT**
  - **R**edundant
  - **O**bsolete
  - **T**rivial
- The question is which data to eliminate?
  - Most enterprise data is never analyzed

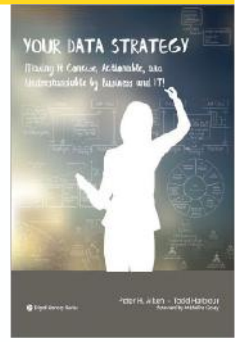
Data Strategy and Governance in Strategic Context



Data Strategy and Governance in Strategic Context



Program Overview



- A data strategy specifies how data assets are to be used to support the organizational strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- A data strategy is necessary for effective data governance
  - Improve your organization's data
  - Improve the way people use their data
  - Improving how people use data to support their organizational strategy
- Effective Data Strategy Prerequisites
  - Lack of organizational readiness
  - Failure to compensate for the lack of data competencies
  - Eliminating the barriers to leveraging data, the seven deadly data sins
- Data Strategy Development Phase II–Iterations
  - Lather, rinse, repeat
  - A balanced approach is required
  - Establish various data value chains
- Q&A

Data Strategy Best Practices

Data Strategy Framework (Part 1)

Business Needs

- Organization Mission
- Strategy & Objectives
- Organizational Structures
- Performance Measures

Business Needs → **Solution**

- Benefits & Success Criteria
- Capability Targets
- Solution Architecture
- Organizational Development

Data Strategy Framework (Part 1)

Business Needs

- Organization Mission
- Strategy & Objectives
- Organizational Structures
- Performance Measures

Current State

- Organizational / Readiness
- Business Processes
- Data Management Practices
- Data Assets
- Technology Assets

Business Needs →

Strategic Data Imperatives

- Business Value Targets
- Capability Targets
- Tactics
- Data Strategy Vision

← Existing Capabilities

Execution ↓

Data Strategy is Implemented in 2 Phases

Data Strategy

What the data assets do to support strategy

Phase I-Prerequisites

1) Prepare for dramatic change and determine how to do the work

2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)

3) Eliminate the Seven Deadly Data Sins

↓

Phase II-Iterations (Lather, Rinse, Repeat)

Data Strategy is Implemented in 2 Phases

Data Strategy

What the data assets do to support strategy

Phase I-Prerequisites

1) Prepare for dramatic change and determine how to do the work

2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)

3) Eliminate the Seven Deadly Data Sins

Phase II-Iterations (Lather, Rinse, Repeat)

CIOs aren't

CDO Job Description

There are more Chief Digital Officers than there are Chief Data Officers

Reporting to senior leadership, the CDO is the data leader responsible for evolving data practices to better support the organizational mission.

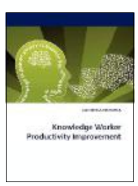
Improving organizational data practices extends the CDO's responsibilities to every knowledge worker in the organization. Empowering knowledge workers with better data practices is the single most important productivity improvement that organizations can make. The CDO is responsible for growing not just an organizational data team but for operationalizing an organization-wide conversation and focus on data innovation, improvement, and value.

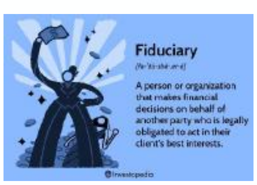
The CDO establishes, fiduciary responsibilities through stewardship, aimed at leveraging data assets and organizational capabilities and creating a climate of data sharing. Some of this can be accomplished by leading the organizational data governance program to effectiveness. The data leader will be required to understand how to appropriately incorporate change management capabilities to the substantive people, process, and ethical challenges that will support the new data focus.

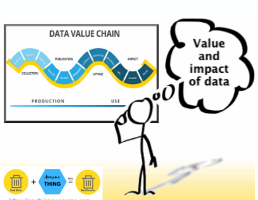
As an organization's sole, non-depletable, non-degrading, non-rivalrous strategic asset, its data has likely been suffering from data debt. The CDO must nurture programs to improve useful subsets of organizational data and simultaneously reduce the impact of data debt. Data volume and debt necessitate prioritization and the CDO must incorporate a strategic approach to improving the value of an organization's data.

For data's true value to become apparent, it needs to be understood as a defined part of the organizational value chain. The CDO is responsible for appropriate aspects of monetization to the organizations data. This requires architecting organizational data requirements in the context of present and future business operations. These requirements identify data products directly supporting business value.

CDO Job Description







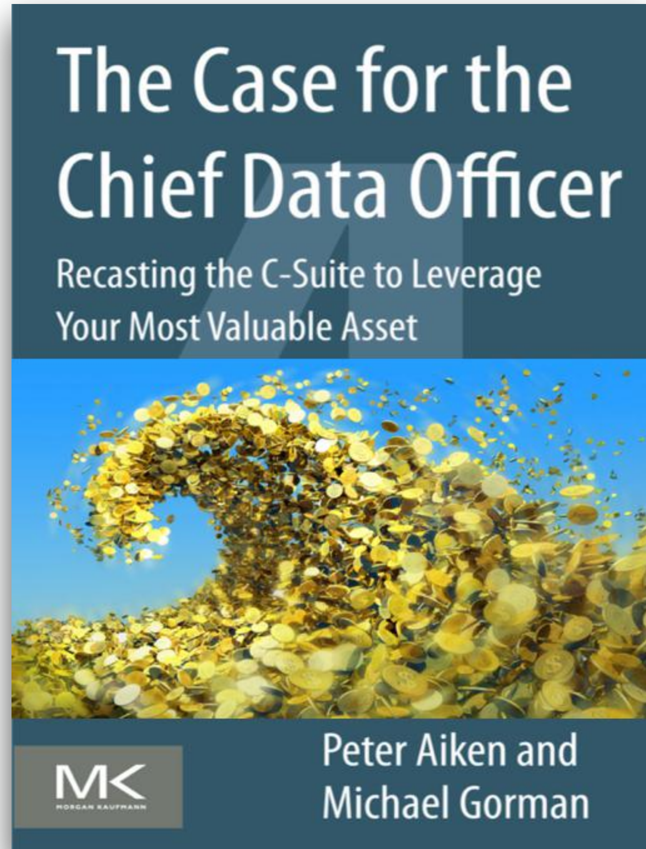
Reporting to senior leadership, the CDO is the data leader responsible for evolving data practices to better support the organizational mission.

Improving organizational data practices extends the CDO's responsibilities to every knowledge worker in the organization. Empowering knowledge workers with better data practices is the single most important productivity improvement that organizations can make. The CDO is responsible for growing not just an organizational data team but for operationalizing an organization-wide conversation and focus on data innovation, improvement, and value.

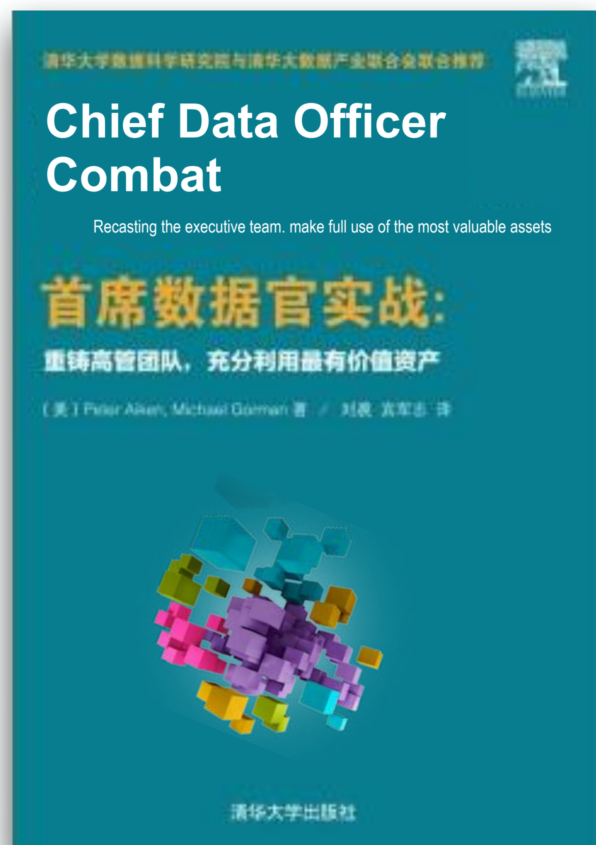
The CDO establishes, fiduciary responsibilities through stewardship, aimed at leveraging data assets and organizational capabilities and creating a climate of data sharing. Some of this can be accomplished by leading the organizational data governance program to effectiveness. The data leader will be required to understand how to appropriately incorporate change management capabilities to the substantive people, process, and ethical challenges that will support the new data focus.

As an organization's sole, non-depletable, non-degrading, non-rivalrous strategic asset, its data has likely been suffering from data debt. The CDO must nurture programs to improve useful subsets of organizational data and simultaneously reduce the impact of data debt. Data volume and debt necessitate prioritization and the CDO must incorporate a strategic approach to improving the value of an organization's data.

For data's true value to become apparent, it needs to be understood as a defined part of the organizational value chain. The CDO is responsible for appropriate aspects of monetization to the organizations data. This requires architecting organizational data requirements in the context of present and future business operations. These requirements identify data products directly supporting business value.



The Case for the Chief Data Officer
Recasting the C-Suite to Leverage Your Most Valuable Asset
MK
MORGAN KAUFMANN
Peter Aiken and Michael Gorman



Chief Data Officer Combat
Recasting the executive team. make full use of the most valuable assets
首席数据官实战：
重铸高管团队，充分利用最有价值资产
清华大学出版社

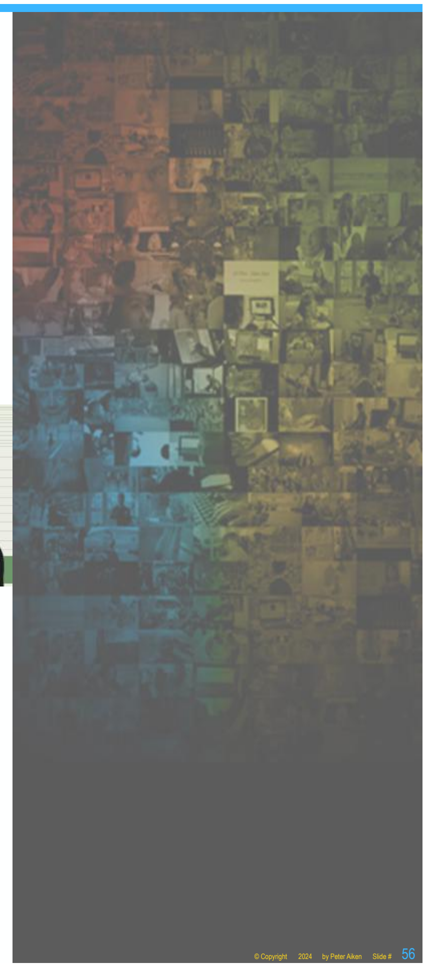
Credit: Image credit: Matt Vickers

Change the status quo!

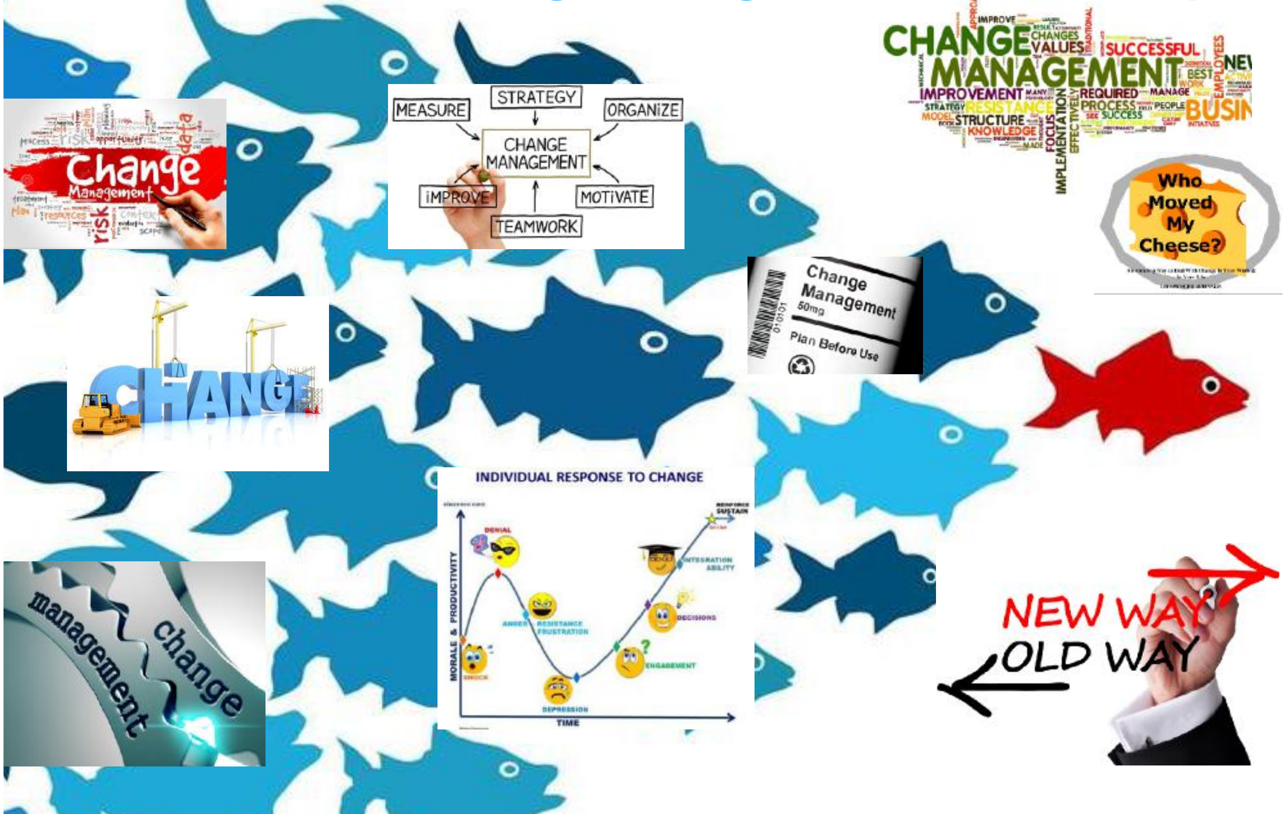
- Keep in mind that the appointment of a CDO typically comes from a high-level decision. In practice, it can trigger an array of problematic reactions within the organization including:
  - Confusion,
  - Uncertainty,
  - Doubt,
  - Resentment and
  - Resistance.

- CDOs need to rise to the challenge of changing the status quo if they expect to lead the business in making data a strategic asset.
  - from *What Chief Data Officers Need to Do to Succeed* by Mario Faria

https://www.forbes.com/sites/gartnergroup/2016/04/11/what-chief-data-officers-need-to-do-to-succeed/#734d53a8434a



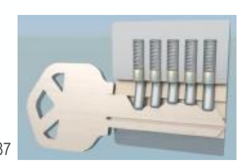
Change Management & Leadership



Diagnosing Organizational Readiness

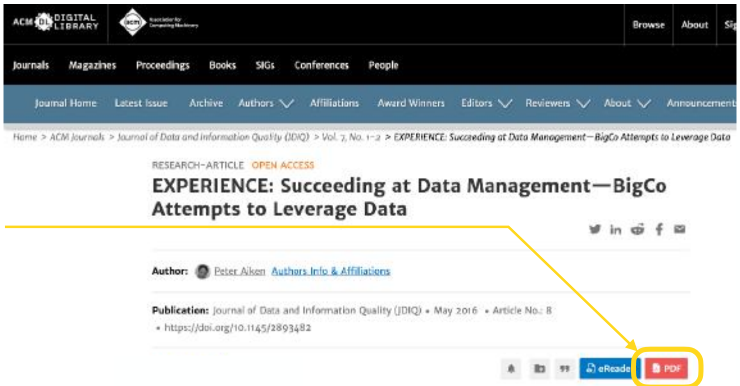
	+	Skills	+	Incentive	+	Resources	+	Action Plan	=	Confusion
Vision	+		+	Incentive	+	Resources	+	Action Plan	=	Anxiety
Vision	+	Skills	+		+	Resources	+	Action Plan	=	Gradual Change
Vision	+	Skills	+	Incentive	+		+	Action Plan	=	Frustration
Vision	+	Skills	+	Incentive	+	Resources	+		=	False Starts
Vision	+	Skills	+	Incentive	+	Resources	+	Action Plan	=	Change

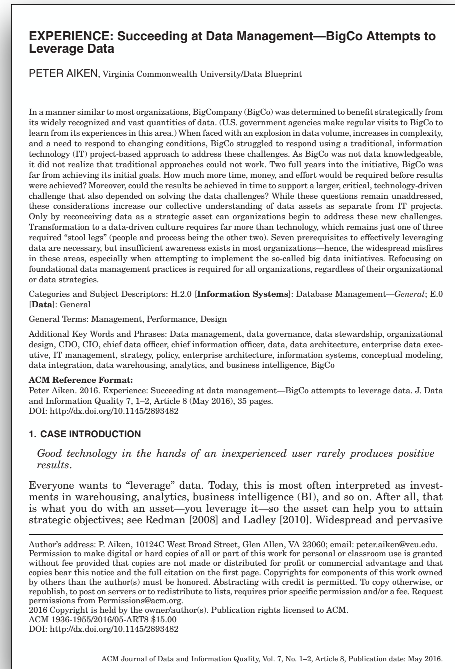
Culture is the biggest impediment to a shift in organizational thinking about data!



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

No cost, no registration case study download





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 unaddressed, 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 "stool 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: peter.aiken@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 republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
2016 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 1936-1955/2016/05-ART8 $15.00
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.

- Download Here!
  - http://dl.acm.org/citation.cfm?doid=2888577.2893482

Data Strategy is Implemented in 2 Phases

Data Strategy

What the data assets do to support strategy

Phase I-Prerequisites

1) Prepare for dramatic change and determine how to do the work

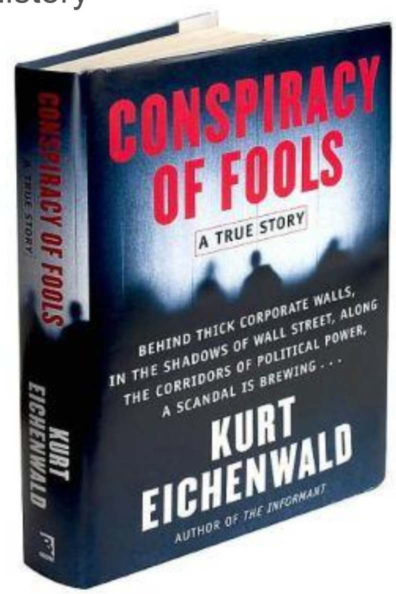
2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)

3) Eliminate the Seven Deadly Data Sins

Phase II-Iterations (Lather, Rinse, Repeat)

Enron

- Fortune named Enron "America's Most Innovative Company" for six consecutive years
- Suffered the largest Chapter 11 bankruptcy in history (up to that time)
- August 2001: $90.00 → $42.00 → $0.26
- Dynegy (several $ billion) attempted rescue
- Enron spends entire amount in 1 week
  - Any person can write a check at Enron for
  - Any amount of money for
  - Any purchase at
  - Any time ...
- Enron goes back to Dynegy for more $?
- Dynegy: What happened to the several $ billion I gave you last week?
- Enron:



http://en.wikipedia.org/wiki/Enron

CFO Necessary Prerequisites/Qualifications

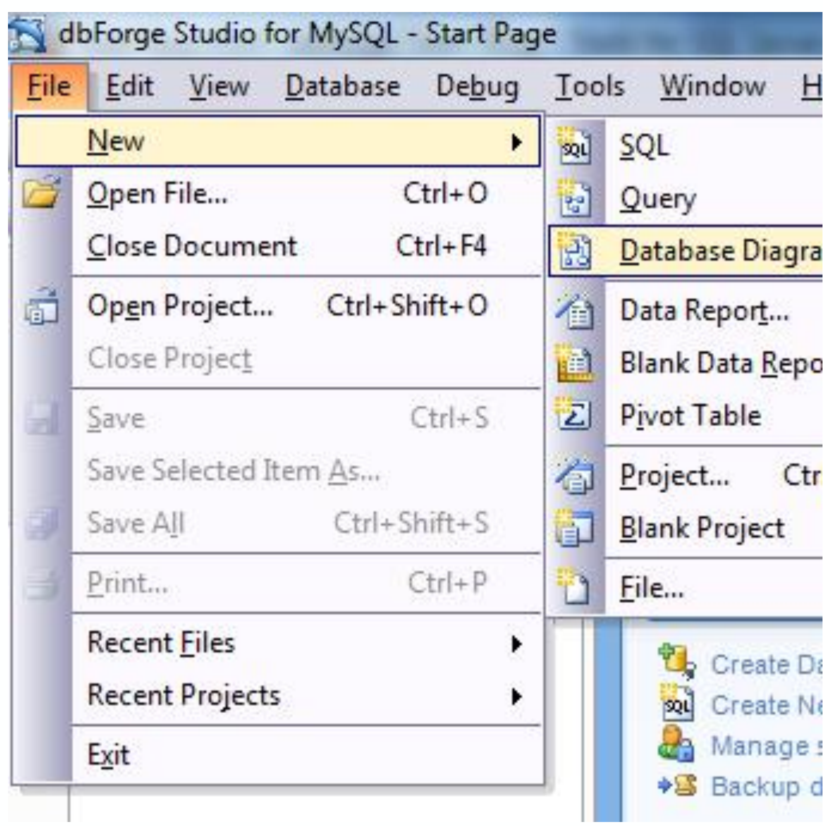
- CPA
- CMA
- Masters of Accountancy
- Other recognized degrees/certifications
- These are necessary but insufficient prerequisites/ qualifications

What do we teach knowledge workers about data?

What percentage of the deal with it daily?

What do we teach IT professionals about data?

- 1 course
  - How to build a new database
- What impressions do IT professionals get from this education?
  - Data is a technical skill that is needed when developing new databases



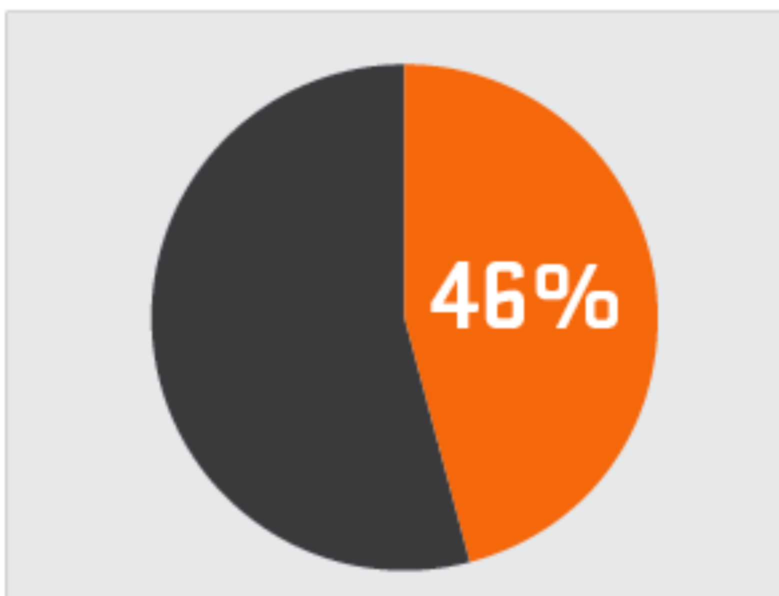
If the only tool you know is a hammer you tend to see every problem as a nail (slightly reworded from Abraham Maslow)

Bad Data Decisions Spiral

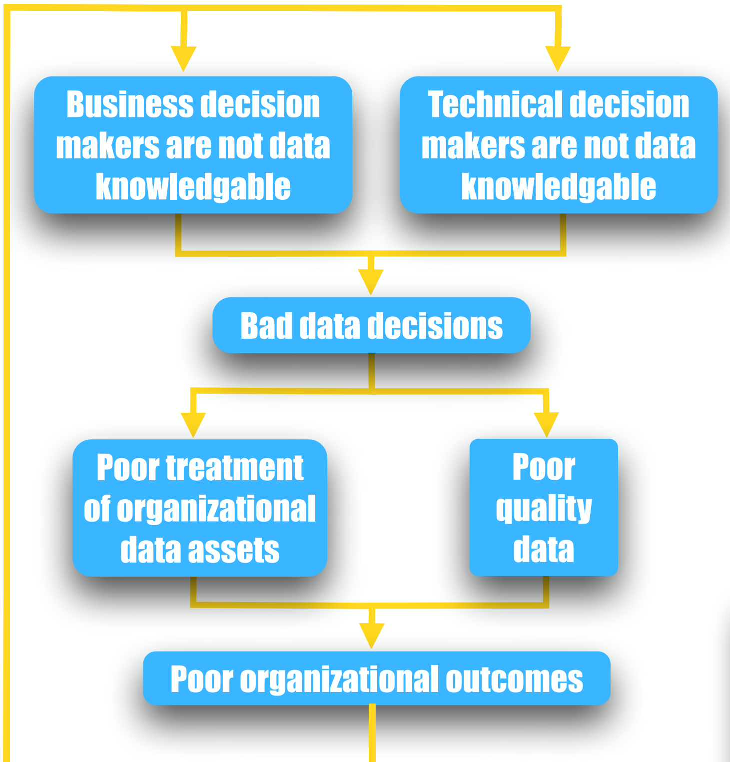
NEWS FLASH!

46% of companies report they made an inaccurate business decision based on bad or outdated data. Bad data leads to bad business decisions. Companies need to be careful that their data is sound – especially when dealing with investors.

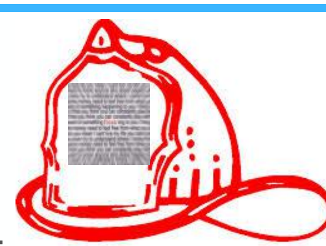
Like Comment Share



Bad Data Decisions Spiral



A Single Focus

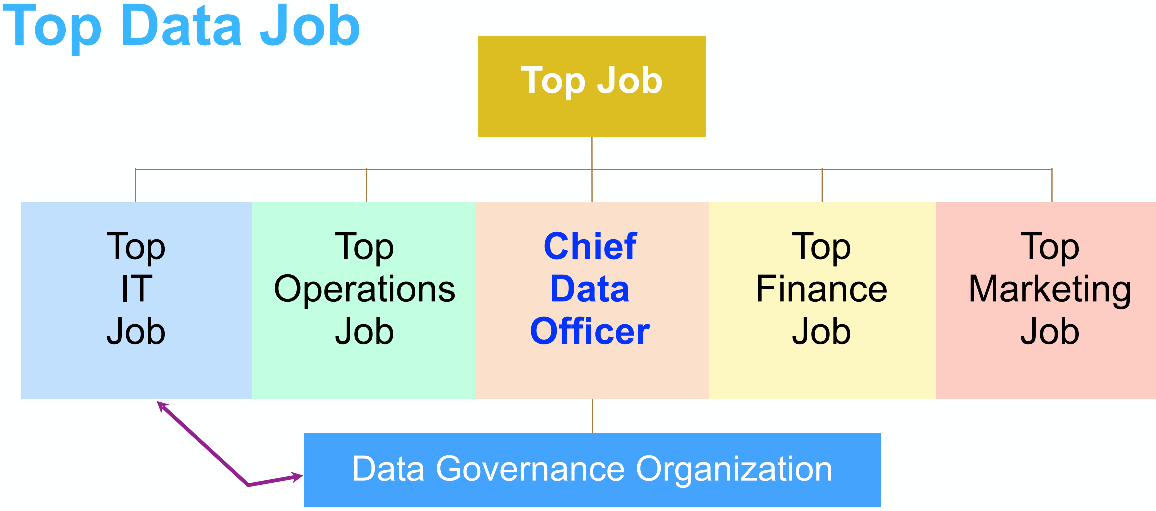


- Chief
  - The head or leader of an organized body of people; the person highest in authority: the chief of police
- Chief Financial Officer (CFO) ← *does not balance books*
  - Individual possessing the knowledge, skills, and abilities to be both the final authority and decision-maker in organizational financial matters
- Chief Risk Officer (CRO) ← *does not test software*
  - Individual possessing the knowledge, skills, and abilities makes decisions and implements risk management
- Chief Medical Officer (CMO) ← *does not perform surgery*
  - Responsible for organizational medical matters. The organization, and the public, has similar expectations for any of chief officer – especially after the Sarbanes-Oxley bill.

[dictionary.com]

Hiring Panels Are Often Challenged to Help

Top Data Job



- Dedicated solely to data asset leveraging
- Unconstrained by an IT project mindset
- Reporting to the business

Data Strategy is Implemented in 2 Phases

Data Strategy

What the data assets do to support strategy

Phase I-Prerequisites

1) Prepare for dramatic change and determine how to do the work

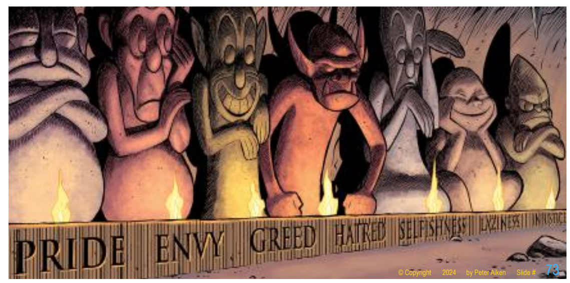
2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)

3) Eliminate the Seven Deadly Data Sins

Phase II-Iterations (Lather, Rinse, Repeat)

Exercising the Seven Deadly Data Sins

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



the Data Doctrine® (V2)

THE DATA DOCTRINE

We are uncovering better ways of developing
IT systems by doing it and helping others do it.
Through this work we have come to value:

data programmes driving IT programs

informed information investing over technology acquisition activities

stable, shared organizational data over IT component evolution

data reuse over the acquisition of new data sources

Inspiration from: theagiledoctrine.org

the Data Doctrine® (V2)

THE DATA DOCTRINE

We are uncovering better ways of developing
IT systems by doing it and helping others do it.
Through this work we have come to value:

data programmes driving IT programs

informed information investing over technology acquisition activities

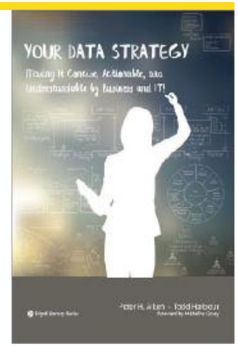
stable, shared organizational data over IT component evolution

data reuse over the acquisition of new data sources

That is, while there is value in the items on
the right, we value the items on the left more.

Source: theagiledoctrine.org

Program Overview



- A data strategy specifies how data assets are to be used to support the organizational strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- A data strategy is necessary for effective data governance
  - Improve your organization's data
  - Improve the way people use their data
  - Improving how people use data to support their organizational strategy
- Effective Data Strategy Prerequisites
  - Lack of organizational readiness
  - Failure to compensate for the lack of data competencies
  - Eliminating the barriers to leveraging data, the seven deadly data sins
- Data Strategy Development Phase II–Iterations
  - Lather, rinse, repeat
  - A balanced approach is required
  - Establish various data value chains
- Q&A

Data Strategy Best Practices

Data Strategy is Implemented in 2 Phases

Data Strategy

Phase I-Prerequisites

1) Prepare for dramatic change and determine how to do the work
2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)
3) Eliminate the Seven Deadly Data Sins

You are here

Phase II-Iterations (Lather, Rinse, Repeat)

1) Identify the primary constraint keeping data from fully supporting strategy
2) Exploit organizational efforts to remove this constraint
3) Subordinate everything else to this exploitation decision
4) Elevate the data constraint
5) Repeat the above steps to address the new constraint

Data Strategy Framework (Part 2)

Business Needs
- Organization Mission
- Strategy & Objectives
- Organizational Structures
- Performance Measures

Current State
- Organizational / Readiness
- Business Processes
- Data Management Practices
- Data Assets
- Technology Assets

Business Needs → **Strategic Data Imperatives** ← Existing Capabilities

Strategic Data Imperatives
- Business Value Targets
- Capability Targets
- Tactics
- Data Strategy Vision

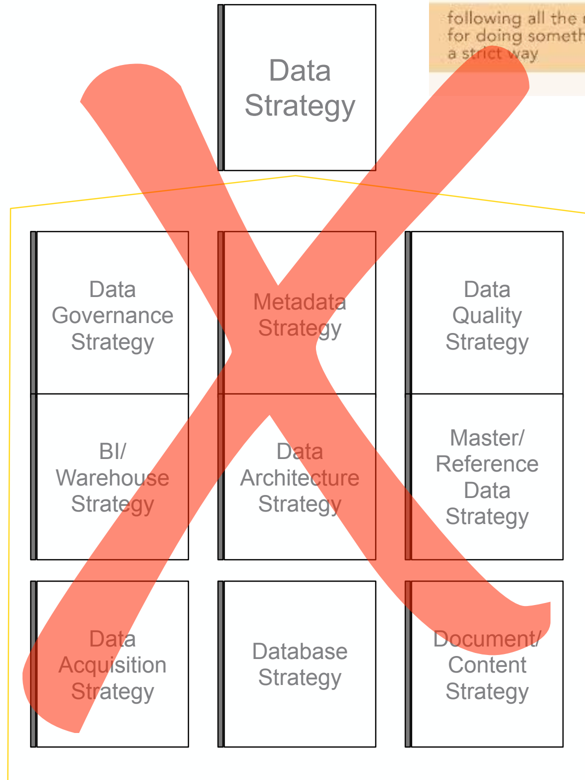
Execution

Road Map
- Leadership & Planning
- Project Dev. & Execution
- Cultural Readiness

Business Value

New Capabilities

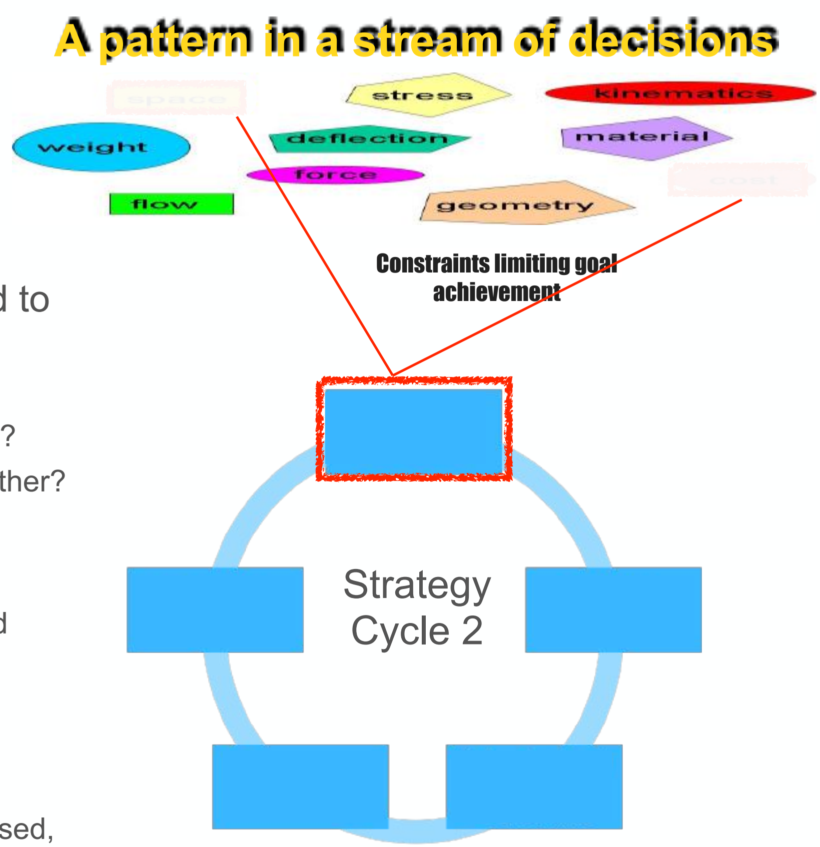
By the (other) Books



Recap

LET'S RECAP

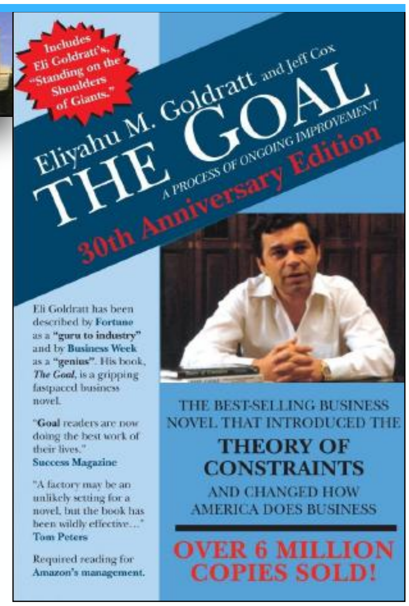
- A data strategy specifies how data assets are to be used to support strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- Strategy evolves periodically
  - As data debt is reduced
  - Determine the most fixable/highest value constraint
  - Cycle through
  - If constraint not addressed, start over



Article Talk

Theory of constraints

From Wikipedia, the free encyclopedia



(TOC)

- A management paradigm that views any manageable system as being limited in achieving more of its goals by a small number of constraints (Eliyahu M. Goldratt)
- There is always at least one constraint, and TOC uses a focusing process to identify the constraint and restructure the rest of the organization to address it

- TOC adopts the common idiom "a chain is no stronger than its weakest link," processes, organizations, etc., are vulnerable because the weakest component can damage or break them or at least adversely affect the outcome

https://en.wikipedia.org/wiki/Theory_of_constraints

Each cycle has an articulated purpose

Theory of Constraints - Generic

Identify the Constraint — Identify the current constraints, the components of the system limiting goal realization

Exploit the Constraint — Make decisions about how best to exploit

Subordinate all non-Constraints — Organizationally–work effectively (not efficiently)

Alleviate the Constraint — Attempt to increase capacity in light of new system demands and non-constraints

Repeat the Process — Repeat until the constraint is eliminated

https://en.wikipedia.org/wiki/Theory_of_constraints

Data Strategy Cycle

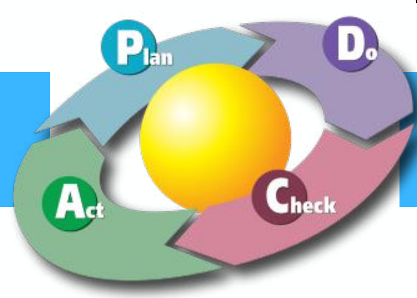
Identify the Constraint — In your analysis of how organization data can best support organizational strategy one thing is blocking you most - identify it!

Exploit the Constraint — Determine how best to exploit the constraint

Subordinate all non-Constraints — Harvest/harness excess capacity of non-constraints

Alleviate the Constraint — Restructure to further increase supporting capacities

Repeat the Process — Repeat until data better supports strategy



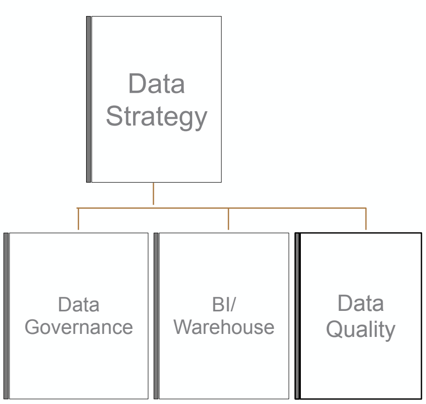
Data Management Body of Knowledge (DM BoK V2)

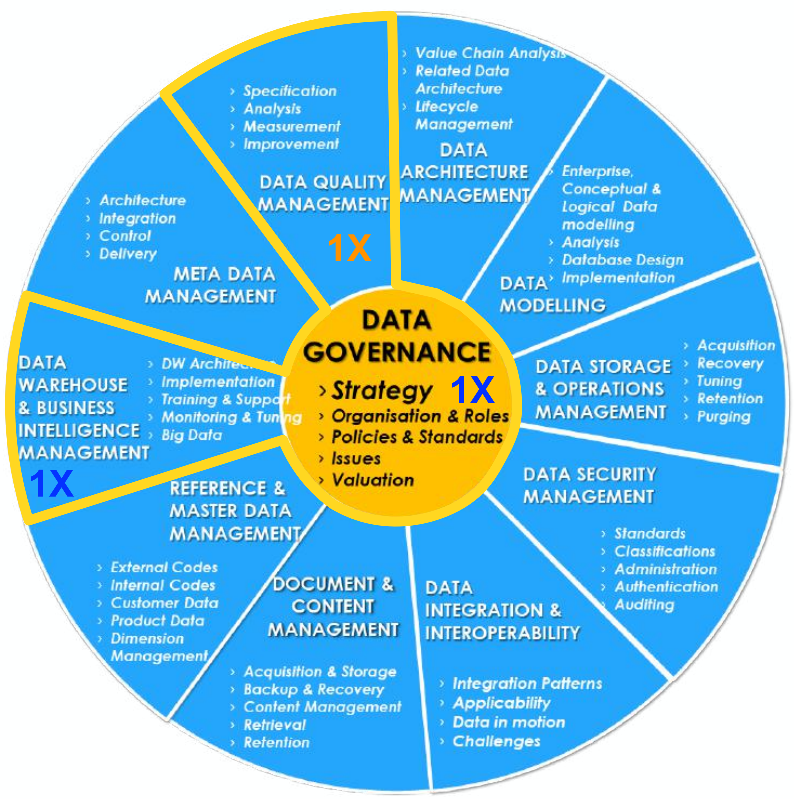
Practice Areas

from *The DAMA Guide to the Data Management Body of Knowledge* 2E © 2017 by DAMA International

Iteration 1

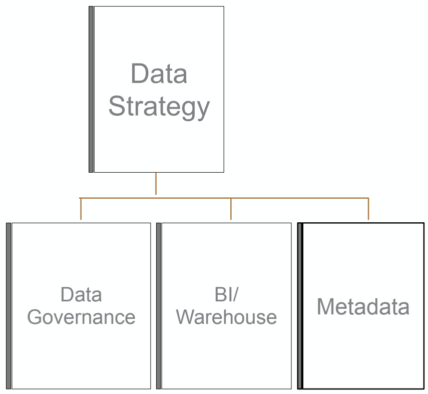
Perfecting operations in 3 data management practice areas





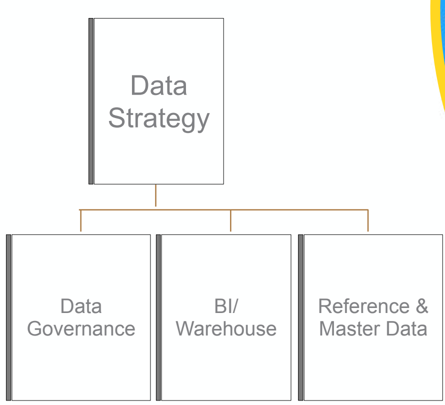
Iteration 2

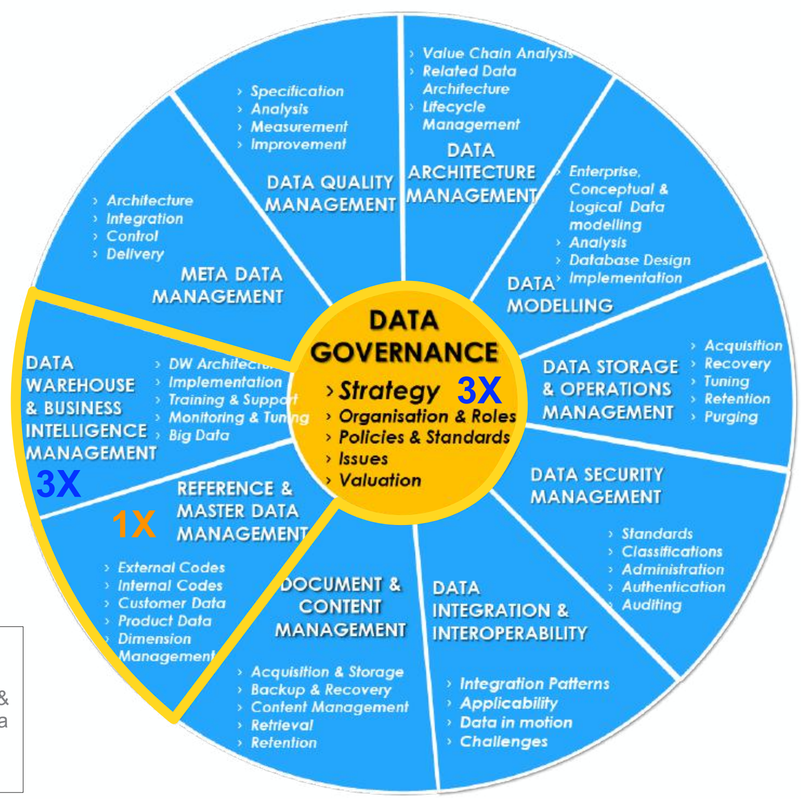
Perfecting operations in 3 data management practice areas



Iteration 3

Perfecting operations in 3 data management practice areas

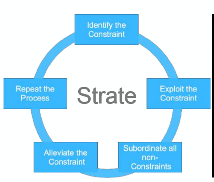


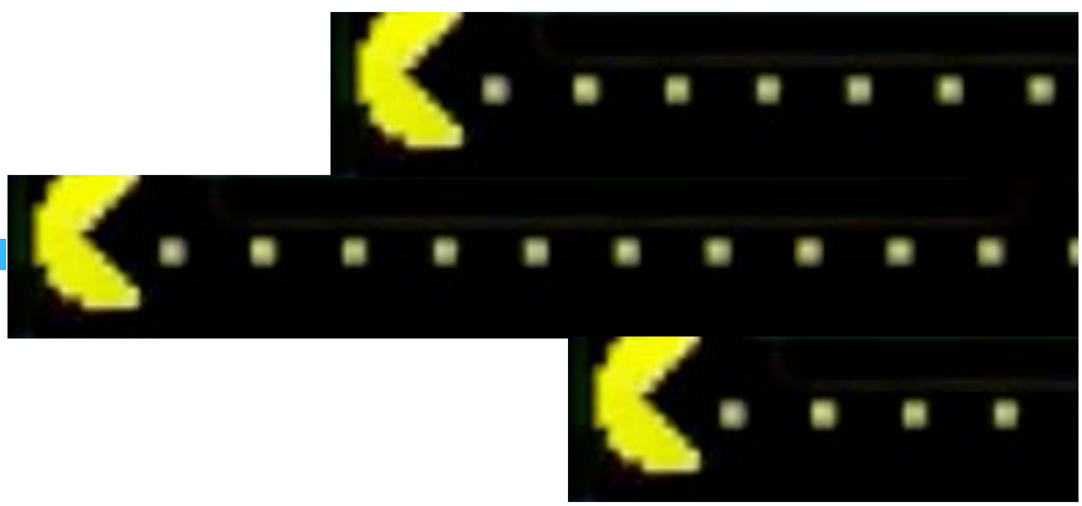


Value Chain

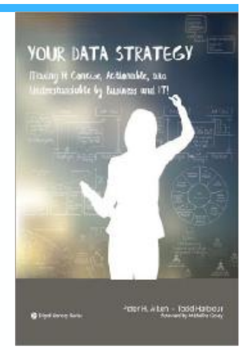
- Consider who knows?
  - Data professionals?
  - Knowledge workers 10x multiplier
  - A sequence of intellectual tasks by which knowledge workers build their employer's unique competitive advantage [1] and/or social and environmental benefit. [https://online.hbs.edu/blog/post/what-is-value-chain-analysis]

Increasing Implementation Effectiveness





Program Overview



- A data strategy specifies how data assets are to be used to support the organizational strategy
  - What is strategy?
  - What is a data strategy?
  - How do they work together?
- A data strategy is necessary for effective data governance
  - Improve your organization's data
  - Improve the way people use their data
  - Improving how people use data to support their organizational strategy
- Effective Data Strategy Prerequisites
  - Lack of organizational readiness
  - Failure to compensate for the lack of data competencies
  - Eliminating the barriers to leveraging data, the seven deadly data sins
- Data Strategy Development Phase II–Iterations
  - Lather, rinse, repeat
  - A balanced approach is required
  - Establish various data value chains
- Q&A

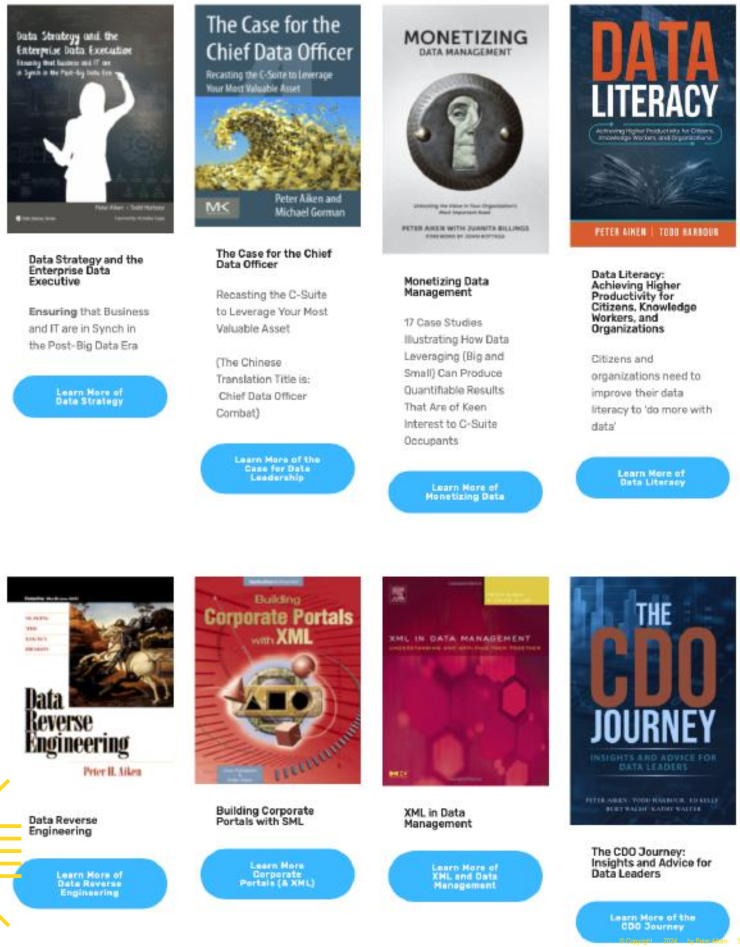
Data Strategy
Plans Are Useless but Planning is Invaluable

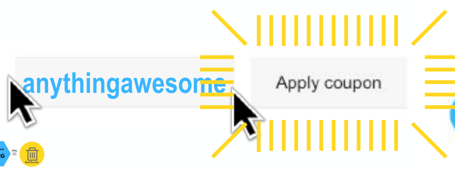
Bottom Line Up Front (BLUF)

- Multi-dozen+ page data strategies are less useful than the process of creating them, especially at first
- Too much time spent writing the perfect plan is accomplished at the expense of the equal effort required to become proficient implementing data strategically
- Cycling through a series of improvements is a better way to think about using data strategically than a grand plan

Event Pricing on Peter's Books

- 20% off directly from the publisher on select titles
- My 'Book Store' @ https://anythingawesome.com/books-overview.html
- Enter the code "anythingawesome" at the Technics bookstore checkout where it says to "Apply Coupon"





Upcoming Events

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

Data Modeling Fundamentals
13 February 2024

The Roles of Data Stewards
12 March 2024

Reference vs Master Data Management
9 April 2024

Brought to you by: DATAVERSITY

Note: In this .pdf, clicking any webinar title opens the registration link

Critical Design Review?

Mentoring?

Executive Data Literacy Training?

Collaboration?

Independent Verification & Validation

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

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

Data Management Maturity Assessment of Public Sector Agencies

Abstract:

To determine how critical data assets are conceptualized and managed in the public sector, we undertake a large-scale empirical study at 15 government agencies. We use the Data Management Maturity (DMM) reference model framework to conduct a systematic multi-level analysis (inter-agency, intra-agency, and cross-case analysis). To aid the comparative assessment of multiple independent agencies, we propose and test the DMM Index. The study addresses many challenges associated with the use of stage models in e-government research. It not only evaluates stage model as a viable framework for the assessment of DMM in the public sector, but also demonstrates how an enterprise-wide systematic assessment may be conducted. The approach presented in the paper can be replicated at other large government entities and private conglomerates. A manager may apply the approach to develop a custom roadmap for data management improvements that align with the organization's business goals.

Keywords: Data management maturity, stage model, DMM model, DMM Index, DMMI, Government Agencies, Public sector.

1. Introduction

Information and communication technologies (ICT) offer great opportunities for government agencies to enhance the effectiveness and efficiency of their internal and external operations. Internally, ICT enables agencies to adapt to changes in governance policies and processes. Externally, it provides rich capabilities that facilitate service convergence and citizen participation. However, providing seamless services and reliable information require more than just scalable and secure technology capabilities. Also required are information strategies that emphasize management, quality, and governance of data, and organizational practices that enfold a strong compliance program, ongoing training, and data sharing practices. Thus, achieving effectiveness and efficiency is not just a matter of capitalizing on ICT capabilities (Gottschalk, 2009), but achieving higher levels of maturity in the management of technology, information, and organizational processes.

The growing importance of data management in government organizations cannot be overstated. As with any organization, data is a critical asset to government agencies. Government organizations collect and maintain different kinds of data related to citizens (Chun, Shulman, Sandoval, & Hovy, 2010). Data ascribes meaning to the information that government agencies wish to share (Chun et al., 2010). Data drives the information discovery needs of citizens and government agencies (Janowski, 2015). It enables information exchange among public agencies and external partners (Gottschalk, 2009). Government agencies integrate and reconcile data from different sources according to the functions and services they provide. Thus, data management is central to achieving operational effectiveness, reducing costs, and improving efficiencies of government services. Agencies recognize the strategic importance of data (Dennis, 2018) and many have developed initiatives to intensify emphasis on data management (VITA, 2017).

Management of data across the various business functions within the government is vital for a variety of reasons. First, data that is undefined and fragmented adds complexity, costs, errors, and inefficiency. Second, data management strategies reduce costs associated with decision making (Kim, Trimi, & Chung, 2014). To facilitate effective decision making, data should be built, stored, nurtured, shared, and managed in a sustainable manner. Third, data management ensures clear agreements around how data is transferred and shared among business partners (Janssen, van der Voort, & Wahyudi, 2017). It helps to improve supplier relations and reduce customer service errors. Fourth, operational functions such as application development, data integration, and reporting are dependent on the strength of the underlying data models (Chen & Zhang, 2014). Last but not least, enterprise-wide data management programs facilitate statutory data security compliance and regulatory reporting (Hiller & Bélanger, 2001). Although e-government literature has discussed the need for improving data management (Linders, 2012), studies examining how government agencies conceptualize data management are sparse.

Just as any other organization, government agencies increasingly seek agile, transparent, effective, and accountable data management practices. Government officials discern the importance of maturity in data management to facilitate high quality, meaningful, and understandable data. Yet, many government agencies do not have any data management or data governance plans (Dennis, 2018). The ability to share data (across departmental, organizational, geographic, and institutional boundaries) and integrate business processes (Gottschalk, 2009) face serious challenges when data management practices are not mature. While there is no single approach to address the complexity of data management, sound data practices are pivotal to achieve organizational efficiency and effectiveness and to facilitate rapid decision-making. For example, when faced with complex legislative challenges, improved decision quality would be reached when decision makers have access to critical inputs from all relevant government agencies. A holistic approach in the provisioning of data related services is thus required to ensure the ability to deliver insights and execute decisions. In addition to modernizing information exchange needed to support decision making (Obama, 2009), establishing maturity in data management would also enable better collaboration and participation between public agencies and external partners. However, there is a paucity of published literature on the assessment of data maturity in government agencies. There is also limited guidance informing how such an assessment may be conducted.

This research targets three objectives. First, the study aims to understand how government agencies conceptualize data management. Specifically, this research seeks to determine how critical data assets are managed in government agencies, and how they support agency-specific goals. To address this research objective, we conducted a rigorous multi-level evaluation of data management practices at 15 government agencies at a State in the east coast of the United States. Second, Information Systems (IS) literature identifies many stage models and frameworks that have been proposed to assess and improve organizational data management. In our study, we use the Data Management Maturity model (DMM, 2014), a widely adopted stage model for the assessment of data management practices. Although the strength of stage models is often construed upon its focus on organizational use of technology (Andersen & Henriksen, 2006), not many researchers agree on the suitability of this ‘evolutionistic’ approach for measuring process capabilities within an organization (King & Kraemer, 1984; Maheshwari & Janssen, 2013).

Through our research we aim to examine whether or not the DMM model represents the purported qualities of a stage model, i.e., descriptive, predictive and testable (Bannister & Connolly, 2015). Third, although published literature highlights the growing importance of measurement and benchmarking within governments and organizations, scholarship on related foundations and methods is somewhat dispersed and inconsistent (Maheshwari & Janssen, 2013). Additionally, to the best of our knowledge, e-government literature has not conducted comparative benchmarking and evaluation of multiple independent government agencies using a single stage model.

The study makes three important contributions. First, although numerous maturity models have been proposed in the past, related research is largely conceptual in nature (Poeppelbuss, Niehaves, Simons, & Becker, 2011). Additionally, research that has empirically validated the various stages in a maturity model is sparse (Solli-Sæther & Gottschalk, 2010). Criticisms abound to the effect that stage theory based studies lack in depth of conceptualization (Bannister & Connolly, 2015) and rely solely on observations (Norris & Lloyd, 2006). We address these research gaps and challenges by conducting a multi-agency data maturity assessment using the DMM model. To the best of our knowledge, a large-scale empirical study involving multiple government agencies using a single stage model has not yet been conducted. Second, the DMM model was developed with the intent of assessing data maturity at a single agency. It lacks mechanisms to compare multiple agencies or business functions within an enterprise. In our research, we develop the DMM Index (DMMI) to effectually compare data management maturity across multiple agencies. We demonstrate how standardization of data maturity scores across various process areas can be achieved using DMMI. Third, our study exemplifies how a multi-level (inter-agency, intra-agency, and cross-case) comparative assessment of data management maturity can be conducted. For practitioners, government agencies, and policy makers, the study presents insights and actionable recommendations that are based on a large enterprise-scale real world assessment. The approach we present in this paper can be replicated by other large government entities and private conglomerates.

The rest of the paper is organized as follows. Section 2 provides background and literature review of stage models and the DMM model. Section 3 describes the research methodology, data, and study setting. The analysis (section 4) examines the multi-level DMM assessment at the 15 government agencies. Section 5 discusses underlying perspectives based on the results of the analysis. Implications for research and practice are presented in section 6. Section 7 summarizes limitations of the study and future research directions, and section 8 concludes the paper.

2. Literature Review

2.1. Stage Models

Stage models have a long history across academic disciplines (Solli-Sæther & Gottschalk, 2010). Foundations of stage models in scholarly literature can be traced to Nolan's (1973, 1979) stage theory and the e-government stage model (Layne & Lee, 2001). Stage models have been proposed for various disciplines. For example, in the field of IS, Galliers and Sutherland (1991) proposed the 'Stages of Growth model' that decomposed maturity into six stages, and Carmel and Agarwal (2006) proposed the IT work offshore stage model. In Entrepreneurship, Scott and

Bruce (1987) proposed the five stage growth model. In marketing, the ‘Product Life Cycle’ concept is a widely used typology to describe the stages a product goes through from its initial conception to its removal from the market (Anderson & Zeithaml, 1984). Consulting firms and private entities have also developed a number of stage models. Lee (2010) conducted a qualitative meta-synthesis of 12 stage models developed and published between the years 2000 and 2010. The meta-synthesis indicated that models vary in perspectives and specific details, borrowed components from other models, and built upon previously developed models with the number of stages ranging mostly from four to six. The research proposed five metaphors for e-government stage models, four of which specifically mention information space, the foundation of which is data management.

In e-government, Layne and Lee (2001) proposed a stage model to describe the different stages of development. They posit four stages of transformation to explain the development of e-government. The four stages (i.e., cataloguing, transaction, vertical integration, and horizontal integration) were proposed based on observations of practices at that time, and the related technological challenges fundamental to supporting basic e-government services. Andersen and Henriksen (2006) proposed a reorientation of the e-government maturity model by focusing on IT applications to improve core activities. They took a critical view of the stage model proposed by Layne and Lee (2001) stating that e-government applications and strategies are a ‘better-safe-than-sorry’ approach at the expense of exploring new areas and dimensions of interaction with end-users. They argued the need for reorienting the stage model by reflecting on the core processes and activities, rather than focusing on the technology-enabled front-end.

2.2. Maturity Models

The growth of organizations over discrete periods of time may be best thought as stages. A review of literature indicates that stage models explain not just stages and patterns of growth of organizations, but also the processes, products, and services. In assessing future growth of the organization, decision makers often start with an assessment of the present situation or status within the company. It is preferred that such an assessment be quantifiable, and not just verbal descriptions of the series of predictable stages of growth (Solli-Sæther & Gottschalk, 2010). What is also of interest to decision makers is an assessment of critical success factors at each stage of growth and its measurement against benchmark (Maheshwari & Janssen, 2013) towards a quantifiable path of evolution.

Maturity models emerged as a means to define the critical success factors at each stage of growth as quantifiable concepts. Published literature frequently interchanges the use of maturity models and stage models. Key distinctions between the two are worthy of observation. First, maturity models provide a clear definition of capabilities and critical success factors at each stage and a means to measure respective capabilities along with benchmark performance at a certain point in time (Andersen & Henriksen, 2006). Maturity models can assist organizations achieve continuous improvement (Iversen, Nielsen, & Norbjerg, 1999) by providing a roadmap with established milestones to assess progress, and steps to be taken for future improvements. They provide a means by which decision makers may gather empirical data as the organization transitions through each stage (Andersen & Henriksen, 2006). It enables decision makers monitor whether the organization is achieving continuous improvement (Iversen et al., 1999).

Over the years, more than a hundred maturity models have been proposed (Becker, Knackstedt, & Pöppelbuß, 2009). In IS research, maturity models spiked so much that Poeppelbuss et al. (2011) identified 76 maturity model research papers published over a 15 year period (1996 to 2010). Their analysis found a decreasing trend in the use of maturity models to study ‘IS development’ (e.g., software development and maintenance, cost-benefit analysis, controls systems, etc.), and an increasing trend in ‘IT and organizations’ (e.g., effectiveness of maturity model to assess and improve organizational capabilities).

Maturity model studies have focused on a variety of topics that include assessing IT management (Becker et al., 2009), improving project management performance (Brookes, Butler, Dey, & Clark, 2014), evaluating operational IT effectiveness (Bradley, Pratt, Byrd, Outlay, & Wynn Jr, 2012), product lifecycle management (Vezzetti, Violante, & Marcolin, 2014), business process management (Röglinger, Pöppelbuß, & Becker, 2012), and leveraging big data (Comuzzi & Patel, 2016). In e-government, maturity models have been proposed to improve interoperability among public and private organizations (Gottschalk, 2009; Pardo, Nam, & Burke, 2012), and guide collaborative public engagement through social media (G. Lee & Kwak, 2012). However, most studies are conceptual proposals of stage models, and their applicability has not been systematically validated across multiple agencies. Within the context of e-government, maturity models provide a means to measure capabilities, benchmark performance at a certain point in time, and assist decision makers to gather empirical data as transitions are made through each stage (Andersen & Henriksen, 2006).

Maturity models have received a fair share of criticism that is attributable to a variety of differing scholarly viewpoints. A vast number of stage models assume that a predictable pattern of growth exists within organizations, and that the patterns unfolding over discrete time periods can be characterized as stages (Gottschalk, 2009). Many stage models also assume that the progression through stages involves a broad range of organizational activities and structures, and that this hierarchical progression is not easily reversible (Gottschalk, 2009). From a prescriptive viewpoint, it is assumed that the later stages are better than the early stages. From a normative viewpoint, there is higher impetus for organizations to achieve higher stages of maturity rather than review performance at a snapshot in time (Andersen & Henriksen, 2006; Poeppelbuss et al., 2011). Other criticisms include the sequential progression between stages (the bottom stage being the initial state having limited capabilities, and the highest stage representing a conception of total maturity) (Becker et al., 2009) and the absolute measures of performance at each stage. However, in reality, the stages may co-exist and evolve simultaneously within the organization. Irrespective of the criticisms, maturity models remain popular owing to their practical value in evaluating organizational capabilities and setting stage for prioritizing actionable improvement measures.

2.3. DMM

Maturity models have traditionally been used by businesses to improve and measure internal business process capabilities. A good example is CMMI, a comprehensive reference model developed by the Carnegie Mellon’s SEI for software management. Over the years, the CMMI reference model has been revised and extended to other areas. For instance, to account for the growing popularity of data management practices, Carnegie Mellon's SEI released the Data Management Maturity (DMM) model in 2014 as a reference model framework of fundamental data management capabilities. It aims to provide organizations with a standard set of best practices to assess data management capabilities.

We chose the DMM model for this study for a variety of reasons. First, it is widely used for the assessment of data management practices within organizations. Second, as a reference framework, the DMM model would enable the systematic multi-level analysis of how government agencies conceptualize data management. Third, using the DMM model for comparative benchmarking of multiple independent government agencies would enable managers at the respective agencies to identify areas of weaknesses in data management practices, as well develop a roadmap for future improvements that align with enterprise business goals. And finally, the study would validate the use of a stage model as a viable framework for assessment of data maturity in the public sector and subsequent replication at other large government entities.

The DMM model comprises of 20 data management process areas and five supporting process areas that are based on the CMMI model (DMM, 2014 p.3) (see Figure 1 and Table 1). The data management process areas are rolled up into five categories – Data Management Strategy, Data Governance, Data Quality, Data Operations, and Platform and Architecture. The best practices for providing support for the implementation of the data management process areas are identified in the Supporting Processes. The lines in Figure 1 represent the interaction among the different categories. An organization's data management maturity is determined based on a methodological assessment of process areas and infrastructure support practices across five levels: Performed, Managed, Defined, Measured, and Optimized (see Table A1 in Appendix for a brief description of the five levels of maturity).

Figure 1. DMM model (DMM, 2014)

DMM Categories	Process Area ID	Process Area Name
Data Management Strategy	1.1	Data Management Strategy
	1.2	Communications
	1.3	Data Management Function
	1.4	Business Case
	1.5	Program Funding
Data Governance	2.1	Governance Management
	2.2	Business Glossary
	2.3	Metadata Management
Data Quality	3.1	Data Quality Strategy
	3.2	Data Profiling
	3.3	Data Quality Assessment
	3.4	Data Cleansing
Data Operations	4.1	Data Requirements
	4.2	Data Lifecycle Management
	4.3	Provider Management
Platform & Architecture	5.1	Architectural Standards
	5.2	Architectural Approach
	5.3	Data Management Platform
	5.4	Data Integration
	5.5	Historical Data, Retention and Archiving
Supporting Processes	6.1	Measurement and Analysis
	6.2	Process Management
	6.3	Process Quality Assurance
	6.4	Risk Management
	6.5	Configuration Management

Table 1. DMM categories and process areas (CMMI, 2016)

3. Research methodology

3.1. Study setting

The study was conducted at one of the States on the east coast of United States. The research utilized an empirical approach to assess data capability and data maturity at 15 government agencies in this State. The State where the review was conducted is ranked in the top third of Gross State Product (GSP) per capita according to the U. S. Department of Commerce. The study utilized the ‘‘key informants’’ approach for data collection (Benlian & Hess, 2011; Oliveira, Thomas, & Espadanal, 2014). The research commenced with a detailed discussion of the study with a senior official at the State level. The senior official identified 15 agencies within the State to participate in the study. As per the senior official, all participating agencies have significant economic impact on the State. The senior official also identified high-ranked managers at each agency who could participate and collaborate with the researchers on the study. The participants were those who were most involved in and knowledgeable of the data management practices at their respective agencies. They included Chief Information Officers and Directors of Data Quality Management. The researchers provided a clear description of project and research objectives to all participants prior to the start of the study. The participation of high-ranked managers who are most familiar with the organization’s data management practices ensured content validity. To reduce self-reporting bias, all participants were given the opportunity to receive the findings from the study and the benchmarking report of how their agency compared against other agencies where the study was conducted. The study was conducted between Sept 2015 and December 2017.

3.2. Data

Data collection was carried out through in-person interviews conducted at each government agency. Data gathering through in-person interviews facilitate in-depth analysis of complex, contemporary, and under-researched activities (Yin, 2009). Interactions with managers at the agencies enabled researchers to understand the data management process in place, as well as examine relevant documentation. Prior to the interviews, researchers gathered and reviewed documents pertinent to data management at each agency. This information formed the basis of the interview questions. Before the in-person interview, a prepared list of questions related to DMM was provided to the interviewees. On average, the interviews lasted two hours. Two members of the research team participated in each interview, one person asked questions, and the other transcribed responses. State policy did not permit audio recording of the interviews. After the interviews were completed, researchers reviewed responses through post-interview discussion and analyzed the transcriptions. Work products such as data strategy plans and documentation that were shared by the interviewees were also reviewed to gain a clear understanding of the data management processes at the agencies.

The interview responses and the work products were scored using a 126-item CMMI DMM survey instrument. The instrument was developed by CMMI with the intention of providing a tool to consistently assess data maturity within organizations. Since the survey comprehensively covers the entire scope of data maturity assessment, we used it without modification for the study. The data collected were responses to the questions on the survey instrument. This helped to ensure that the data conformed to the themes in the DMM model. To ensure reliability of scoring, a copy of the assessment was sent to the manager to verify that it accurately reflected the maturity of data management processes at the agency. The attainment of functional practices was ranked as none, partial, or full based on the evaluation of supporting evidence provided by the interviewees. This reduced the subjectivity of the evaluation. For example, to assess the functional practice survey item "A data management strategy representing an organization-wide scope is established, approved, promulgated, and maintained," the interviewees provided objective evidence of attainment that included examples of work products such as a data management strategy, a list of data management objectives and priorities, data management policies, stakeholder participation and approval documents, data management program scope documentation (e.g., subject areas, business areas, key data elements, key disciplines, etc.), data management strategy sequence plans, data management program metrics, program cost-benefit analysis results, data management program reviews, and data management strategy dashboards.

DMM developed maturity benchmarks based on survey responses from past assessments. Our scoring of the organization's maturity levels relied on the DMM's benchmark to calibrate the survey responses. Maturity levels for each process area were determined based on the aggregation of the scores using a data maturity scale of 1 to 5, where 1 indicated that a process was Performed (lowest level) and 5 indicated an Optimized process (highest level). While the raw scores were maintained at two decimal precision, we rounded the raw score to the whole numbers to determine the associated maturity levels. We adopted this approach for two reasons.

First, we retained the raw scores at two decimal precision so that we could determine how close the agencies were to achieving the next maturity level. Second, we assigned a maturity level by rounding the raw scores to the pre-defined DMM maturity levels. Doing so enabled us to consistently map the maturity levels of agencies to the linear sequence of stages generally used by stage models.

4. Analysis

The DMM reference model framework was developed as tool to evaluate data management practices at an organization (DMM, 2014). It serves to provide a benchmark of performance against which all future DMM initiatives may be measured. DMM depends on the levels at which the data management process areas and supporting process areas are performed, managed, defined, measured, and optimized.

We conducted analysis at three levels. First, we looked at the trends at the enterprise level (i.e., the State). We did so to understand the overall landscape of data management maturity across all 15 State agencies. The analysis was essential to gain an understanding of the process areas that were most mature across the agencies. Next, we conducted an intra-agency analysis to assess maturity of DMM categories within agencies. The objective of this analysis was to shed light on the most common maturity levels across the DMM categories and how they compared against the Supporting Processes. Lastly, we conducted a cross-case analysis by examining similarities and differences between two agencies that have contrasting data maturity assessment scores. The purpose of the cross-case analysis was to investigate if and why there exist differences in the levels of data management maturity among agencies within the State.

4.1. Inter-agency analysis

Although the DMM model does not specify a recommended level of maturity, we chose level 3 (Defined) as the target for our assessments. We chose level 3 for three reasons. First, CMMI (2002) recommends level 3 for capability maturity assessment using CMM. Since CMM serves as the blueprint for the DMM model, using level 3 as a target for our assessment is justifiable. Second, past studies have been restricted to maturity level 3 or lower, as a vast majority of organizations seek capability maturity at level 3. Only a few large organizations have obtained CMM level 5 status and small organizations rarely move beyond level 3 (Swinarski, Parente, & Kishore, 2012). And third, the United States Department of Defense directive 5000 mandates that contractors must be CMMI level 3 certified or achieve CMMI level 3 certification within 18 months of contract award. Thus, achieving level 3 is generally considered indicative that the organization has the wherewithal to successfully develop and implement a DMM program.

The raw process area scores for all 15 agencies are shown in Table 2. Our analysis shows that two agencies, Motor Vehicles and Education, have a majority of process area scores approaching level 3 (Defined) maturity. Additionally, five agencies (Behavioral Health & Developmental Services, Taxation, Aging & Rehab Services, Health, and Conservation & Recreation) have a majority of process area scores approaching level 2 (Managed) maturity. The remaining eight agencies have process area scores mostly at level 1 (Performed). Notably, all process areas of Behavioral Health & Developmental Services achieved *Managed* or above.

DMM Categories	Process Area Name	Maximum	Minimum	Mode	Average	Median	Motor Vehicles	Education	Behavioral Health and Developmental Services	Taxation	Aging and Rehab Services	Health	Conservation & Recreation	Treasury	Corrections	Transportation	Social Services	Health Professions	Housing & Community Development	Veterans Services	Medical Assistance Services
Data Management Strategy	Data Management Strategy	2.75	0.50	0.50	1.13	1.00	2.75	1.75	1.75	1.75	0.50	1.25	1.75	0.50	1.00	0.50	1.00	0.50	0.50	0.50	1.00
	Communications	2.75	0.50	1.50	1.62	1.50	2.75	2.50	2.75	2.00	2.50	1.50	1.75	0.50	1.50	1.50	1.00	1.00	1.50	1.00	0.50
	Data Management Function	2.75	0.50	0.50	1.28	1.00	2.75	2.75	1.75	1.75	1.75	1.75	1.00	0.50	0.50	1.25	0.50	1.00	0.50	1.00	0.50
	Business Case	5.00	0.25	0.50	1.38	0.75	5.00	0.50	1.75	0.75	5.00	1.25	2.75	0.50	0.25	0.75	0.50	0.25	0.50	0.75	0.25
	Program Funding	5.00	0.00	0.00	1.15	0.50	3.50	1.75	2.50	0.00	5.00	1.75	0.00	0.00	0.00	1.75	0.50	0.00	0.00	0.00	0.50
Data Governance	Governance Management	2.75	0.00	1.00	1.08	1.00	1.25	1.75	2.75	1.00	2.75	2.50	0.75	0.25	0.50	1.00	1.00	0.50	0.00	0.00	0.25
	Business Glossary	2.75	0.00	0.00	1.07	0.75	2.75	2.25	1.75	1.50	0.75	2.25	2.00	0.25	0.50	0.25	1.25	0.50	0.00	0.00	0.00
	Metadata Management	3.75	0.00	1.75	1.45	1.50	3.50	1.75	1.75	2.00	3.75	2.75	1.00	1.50	1.00	1.75	0.50	0.00	0.00	0.50	0.00
Data Quality	Data Quality Strategy	3.75	0.00	0.75	1.20	0.75	3.75	0.75	2.75	1.75	0.75	2.50	1.75	0.50	1.00	0.50	0.50	0.75	0.00	0.75	0.00
	Data Profiling	3.75	0.00	0.00	1.02	0.50	1.75	3.75	1.75	0.50	1.75	2.50	0.00	1.25	1.00	0.50	0.00	0.50	0.00	0.00	0.00
	Data Quality Assessment	2.25	0.00	0.50	1.02	1.00	1.25	1.75	1.75	0.50	2.25	2.25	1.50	0.50	1.00	0.50	1.00	0.00	0.00	1.00	0.00
	Data Cleansing	2.75	0.00	0.00	1.23	1.25	1.00	2.75	2.75	1.75	0.00	1.75	1.50	1.25	2.00	0.50	1.50	0.50	1.25	0.00	0.00
Data Operations	Data Requirements	2.75	0.00	0.75	1.13	0.75	0.75	0.75	2.75	1.75	2.50	1.75	2.00	0.75	1.50	0.25	0.50	0.75	0.75	0.25	0.00
	Data Lifecycle Management	2.75	0.00	0.75	0.95	0.75	2.75	1.75	1.75	1.75	0.75	1.75	0.00	0.75	0.50	0.75	0.50	0.75	0.25	0.00	0.25
	Provider Management	2.75	0.25	0.75	1.32	1.00	2.75	2.75	1.75	1.75	0.75	1.75	1.75	0.75	0.75	1.75	0.50	0.75	0.75	1.00	0.25
Platform & Architecture	Architectural Standards	2.75	0.00	1.75	1.30	1.25	1.50	2.75	1.75	1.75	1.75	1.75	1.25	2.75	1.00	0.50	0.50	0.00	1.25	1.00	0.00
	Architectural Approach	2.75	0.25	1.75	1.18	0.75	1.75	2.75	1.75	2.75	1.75	0.25	0.25	1.75	1.50	0.75	0.50	0.25	0.75	0.25	0.75
	Data Management Platform	4.75	0.00	0.50	1.68	1.50	3.50	2.75	2.75	1.75	4.75	1.25	1.75	2.75	0.50	1.50	0.50	0.50	0.00	0.50	0.50
	Data Integration	3.75	0.00	0.00	1.22	1.25	2.50	3.75	1.75	2.50	0.50	1.50	1.25	1.50	0.00	1.50	0.50	1.00	0.00	0.00	0.00
	Historical Data, Retention and Archiving	3.50	0.50	1.75	1.60	1.75	2.75	2.75	1.75	3.50	1.75	1.75	1.75	2.00	0.50	0.50	0.75	0.75	1.75	1.25	0.50
Supporting Processes	Measurement and Analysis	3.25	0.00	0.00	0.73	0.50	0.00	2.75	1.75	0.50	3.25	1.25	0.00	0.50	0.50	0.00	0.00	0.50	0.00	0.00	0.00
	Process Management	2.50	0.00	0.50	0.93	0.75	0.75	2.50	1.75	0.75	0.75	1.25	1.75	1.00	1.00	0.50	0.50	0.50	0.00	0.50	0.50
	Process Quality Assurance	5.00	0.00	0.50	1.70	1.25	5.00	3.50	1.50	2.25	3.50	3.00	1.00	1.75	0.50	0.50	0.50	0.00	1.25	0.00	1.25
	Risk Management	5.00	0.00	0.50	2.38	2.50	2.50	2.75	2.75	5.00	5.00	1.00	2.25	5.00	3.00	0.50	0.50	0.50	4.50	0.50	0.00
	Configuration Management	4.75	0.00	0.00	1.95	2.00	2.50	2.75	3.50	4.75	4.75	0.75	2.00	4.00	2.00	0.50	1.50	0.00	0.00	0.25	0.00
Mode							2.75	2.75	1.75	1.75	1.75	1.75	1.75	0.5	0.5	0.5	0.5	0.5	0	0	0
Level							Defined		Managed					Performed							

Table 2. Process Area Scores for Agencies

Note: Agencies are ordered based on the number of process areas for which they have a data management maturity score of 1.5 or higher. Agency names highlighted in dark grey are closer to level 3 (*Defined*), and agencies in light grey are closer to level 2 (*Managed*). Cells with scores 1.5 or greater are shaded.

While the DMM model provides an approach for the assessment of a single agency, it does not provide a mechanism for the comparative assessment of multiple independent entities within a larger enterprise. A State government with many agencies is one such example. We propose the DMMI to address this limitation. By standardizing data maturity scores, DMMI provides a relative ranking of agencies across all six DMM categories. The heuristics to calculate the DMMI is below:

1. For each category:
    a. Sum the process area maturities within the category.
    b. Force-rank the entities based on the results of Step 1.a.
2. For each agency:
    a. Sum the rankings for the six categories, A_s (the lower the sum, the higher the ranking).
3. For the enterprise:
    a. Force-rank the agencies based on the results of Step 2.a.
    b. Calculate the DMMI as follows: $(1\text{-}A_s / T_r) / (1 - 1/z)$
        where, T_r is the total number of possible rankings, and z is the total number of agencies in the enterprise.
    c. Enterprise agencies with the higher index are those with the higher data management maturities.

For our analysis, A_s represents the rankings for the six DMM categories for the agency, $T_r = 90$ (total number of categories x the number of agencies), and z = 15 (the number of agencies).

	Agency	**Sum of Rankings (*As*)**	**DMMI**
1	Education	16	0.88
2	Motor Vehicles	16	0.88
3	Behavioral Health & Developmental Services	22	0.81
4	Aging & Rehabilitative Services	23	0.80
5	Health	29	0.73
6	Taxation	30	0.71
7	Conservation & Recreation	43	0.56
8	Treasury	52	0.45
9	Corrections	54	0.43
10	Transportation	57	0.39
11	Social Services	65	0.30
12	Housing & Community Development	73	0.20
13	Health Professions	76	0.17
14	Veterans Services	78	0.14
15	Medical Assistance Services	86	0.05

Table 3. Inter-agency DMMI

Table 3 shows the relative ranking of all agencies based on the DMMI. The analysis indicated that both Education and Motor Vehicles attained the highest maturity index (0.88), followed by Behavioral Health & Developmental Service (0.81) and Aging & Rehabilitation (0.80). Together, they represented the top fifth of the index. Similarly, Housing & Community Development (0.20), Health Professions (0.17), Veterans Services (0.14), and Medical Assistance Services (0.05) fell in the bottom fifth of the index. The remaining seven agencies occupy the middle range. Our analysis thus provides a forced-ranking data maturity index for all agencies using a standardized benchmark.

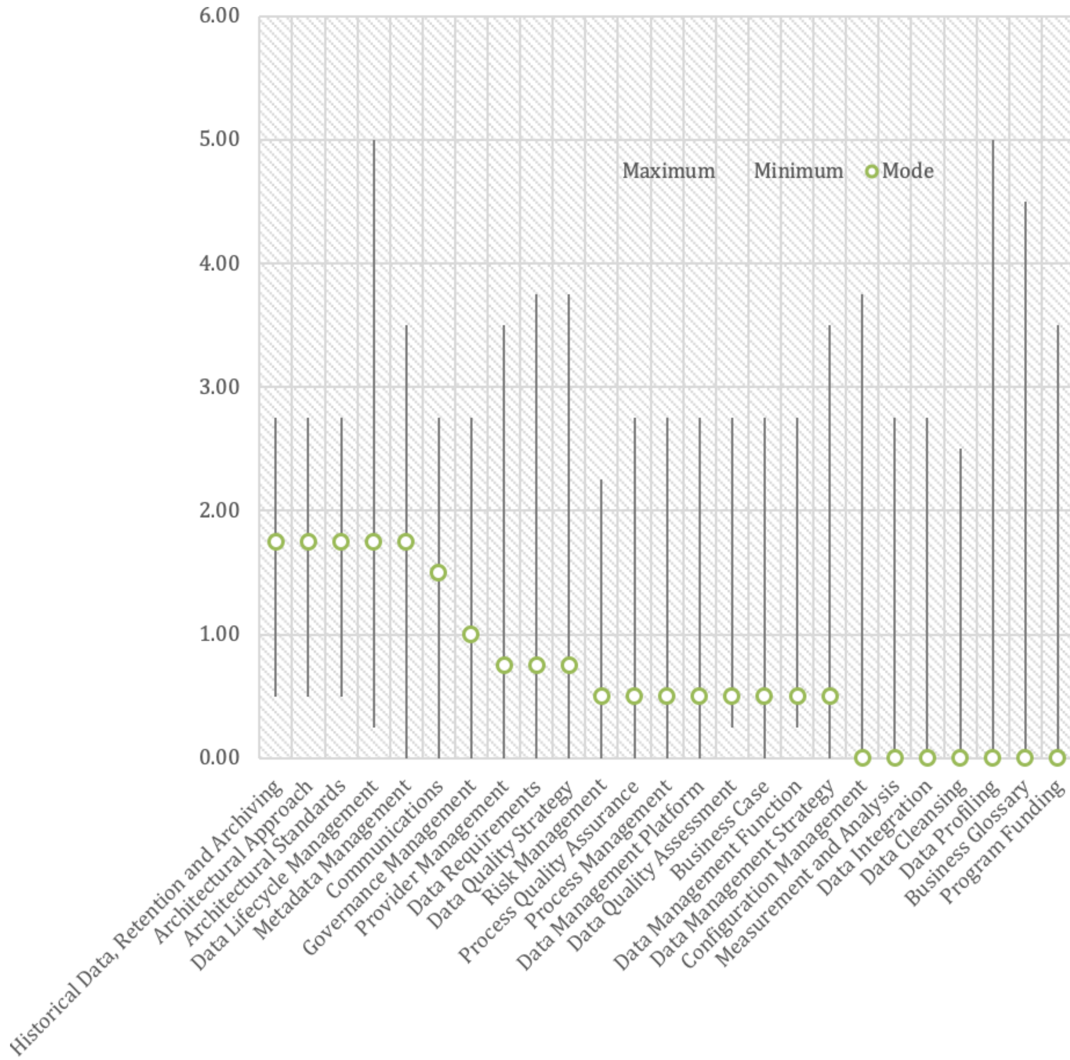


Figure 2. Process area maturity scores

Note: The process areas are ordered based on the most frequently occurring (mode) maturity level across all agencies. Process areas mode values shown in the figure are in the descending order.

To assess the process areas that were most mature across all agencies in the enterprise, we grouped process areas by the maturity level (see Figure 2). Five process areas (Historical Data, Retention, & Archiving, Architectural Approach, Architectural Standards, Data Lifecycle Management, Metadata Management, and Communications) had level 2 (*Managed*) maturity (mode $\geq$ 1.5). For all agencies, the remaining 19 process areas were at level 1 (*Performed*) maturity.

4.2. Intra-agency analysis

To assess category maturity within an agency, we first calculated the mode of the process areas within each category (Table 4). The mode indicates the most common level of maturity across the categories.

DMM Categories	Agencies														
	Motor Vehicles	Education	Behavioral Health and Developmental Services	Taxation	Aging and Rehab Services	Health	Conservation & Recreation	Treasury	Corrections	Transportation	Social Services	Health Professions	Housing and Community Development	Veterans Services	Medical Assistance Services
Data Management Strategy	3	2	2	2	5	2	2	1	1	1	1	1	1	1	1
Data Governance	N/A	2	2	2	N/A	3	1	0	1	N/A	1	1	0	0	0
Data Quality	1	N/A	3	2	2	3	2	1	1	1	1	1	0	1	0
Data Operations	3	N/A	2	2	1	2	2	1	1	N/A	1	1	1	0	0
Platform & Architecture	2	3	2	2	2	2	1	2	1	1	1	1	1	1	1
Supporting Processes	3	3	2	1	5	1	2	1	1	1	1	1	0	0	0

Table 4. Intra-agency category comparison

Note: The values are modes of the process area maturity levels within the category. An N/A indicates that, within a given category none of the process area maturities matched, and hence the mode could not be determined.

For seven agencies the mode for Supporting Processes matched the majority data management category modes. They are represented by the dark grey shaded areas in the table. For example, for Behavioral Health and Developmental Services, the maturity of the Supporting Processes are at level 2, which matches the majority of the data management category maturities (also at level 2). Even though Medical Assistance Services met the criteria for this group, we did not include it because its maturity levels were too low.

For four agencies, the maturity of the Supporting Processes was one level below the majority of data management category modes. They are represented by the light grey shaded areas in Table 4. For example, for Taxation, the maturity of the Supporting Processes (level 1) was one level lower than the majority of the data management category maturities (level 2). For two of the remaining three agencies (i.e., Motor Vehicle and Education), the Supporting Processes maturity was at level 3 (*Defined*). However, for these two agencies the DMM category maturities were not consistent. Additionally, for Aging and Rehab Services, Supporting Processes was at the highest level of maturity (level 5 – *optimized)*. However, the DMM categories were not at a consistent level of maturity.

4.3. Cross-case analysis

Cross-case analysis examines similarities and differences across cases, where the unit of analysis is a case (Khan & VanWynsberghe, 2008; Mathison, 2005). The unit may be an individual, group, place, or organization. By comparing accumulated knowledge from discrete cases, cross-case analysis enables researchers to gain insights and produce knowledge that is broader in scope than a single case. It provides opportunities for researchers to gather evidence and construct explanations as to why one case may be similar or different from another case (Khan & VanWynsberghe, 2008).

For an in-depth cross-case analysis we chose two agencies (Department of Motor Vehicles and Department of Transportation) that are highly significant to the State but have contrasting data maturity assessment scores. The process area scores for the two agencies are shown in Table 5. The table also shows the difference of process area score between the agencies, as well as the maximum, minimum, and mode of process area scores for all 15 agencies. They are provided as a frame of reference.

DMM Categories	Process Area Name	Motor Vehicle	Transportation	Diff	Maximum	Minimum	Mode
Data Management Strategy	Data Management Strategy	2.75	0.50	2.25	2.75	0.50	0.50
	Communications	2.75	1.50	1.25	2.75	0.50	1.50
	Data Management Function	2.75	1.25	1.50	2.75	0.50	0.50
	Business Case	5.00	0.75	4.25	5.00	0.25	0.50
	Program Funding	3.50	1.75	1.75	5.00	0.00	0.00
Data Governance	Governance Management	1.25	1.00	0.25	2.75	0.00	1.00
	Business Glossary	2.75	0.25	2.50	2.75	0.00	0.00
	Metadata Management	3.50	1.75	1.75	3.75	0.00	1.75
Data Quality	Data Quality Strategy	3.75	0.50	3.25	3.75	0.00	0.75
	Data Profiling	1.75	0.50	1.25	3.75	0.00	0.00
	Data Quality Assessment	1.25	0.50	0.75	2.25	0.00	0.50
	Data Cleansing	1.00	0.50	0.50	2.75	0.00	0.00
Data Operations	Data Requirements	0.75	0.25	0.50	2.75	0.00	0.75
	Data Lifecycle Management	2.75	0.75	2.00	2.75	0.00	1.75
	Provider Management	2.75	1.75	1.00	2.75	0.25	0.75
Platform & Architecture	Architectural Standards	1.50	0.50	1.00	2.75	0.00	1.75
	Architectural Approach	1.75	0.75	1.00	2.75	0.25	1.75
	Data Management Platform	3.50	1.50	2.00	4.75	0.00	0.50
	Data Integration	2.50	1.50	1.00	3.75	0.00	0.00
	Historical Data, Retention and Archiving	2.75	0.50	2.25	3.50	0.50	1.75
Supporting Processes	Measurement and Analysis	0.00	0.00	0.00	3.25	0.00	0.00
	Process Management	0.75	0.50	0.25	2.50	0.00	0.50
	Process Quality Assurance	5.00	0.50	4.50	5.00	0.00	0.50
	Risk Management	2.50	0.50	2.00	5.00	0.00	0.50
	Configuration Management	2.50	0.50	2.00	4.75	0.00	0.00
Mode		2.75	0.50				

Table 5. Cross-case analysis scores for DMV and DOT

Note: Diff column shows the difference of scores between Motor Vehicle and Transportation. Maximum, minimum, and mode are process area scores. They are provided as a reference for the cross-case analysis.

The first agency, Department of Motor Vehicles (DMV) is at the *Defined* data maturity level (mode = 2.75). The agency has an operating budget of $0.26 billion for 2016-18 and represents 0.50% of the State's overall budget. It has 2000 employees, and its budget falls within the top 20% of the total operating budgets for the State. The second agency, Department of

Transportation (DOT) is at Performed data maturity level (mode = 0.50). The agency has an operating budget of $5.8 billion for 2016-18 and represents 11% of the State's overall budget. The agency employs 7800 positions, and its budget falls within the top 2% of the total operating budgets for the State.

Our analysis indicated that even though the DMV has a much smaller operating budget than DOT, its data maturity assessment was two maturity levels higher than the latter. A plausible explanation for the higher DM assessment score is that DMV manages and controls sensitive personal information that requires a greater level of confidentiality, integrity, and availability of data. This is not so critical for DOT.

5. Discussion

A DMM assessment provides organizations with empirical data required to improve the maturity of data management strategies, data governance, data quality, and data operations. Improving maturity across these areas of data management within the enterprise offers a pathway to increased collaboration and knowledge sharing (Janssen et al., 2017), elevated organizational performance (DMM, 2014), and improved decision quality (Ghasemaghaei, 2019). For government agencies, attaining greater data management maturity would ensure that decision makers have access to critical inputs needed to analyze complex problems and improve decision quality. For example, data sharing and collaboration among the taxation, education, and health departments would provide a more complete perspective on low-income individuals who qualify and require focused government support.

Our study sought to investigate how government agencies conceptualized data management. By conducting a systematic assessment of DMM at 15 agencies our goal also was to determine characteristics that distinguished data management maturity levels across agencies. The intra-agency assessment aimed to understand maturity levels of DMM categories and Supporting Processes, as well as develop actionable recommendations. We conducted the cross-case analysis to determine whether any transformational steps differentiated the data management maturity levels of the two agencies.

5.1. Inter-agency assessment

An effective approach to conduct an inter-agency comparison would be to use a single aggregate maturity score that is based on the DMM and Supporting Processes categories. Such a score would be beneficial to assess the current state of data management practice in an enterprise setting. However, the DMM reference model framework (DMM, 2014) does not provide a mechanism for the comparative assessment of the maturity of multiple independent entities within a larger enterprise. We proposed the DMMI to fill this gap. For example, our analysis identified Motor Vehicles and Education as leaders in the State with a DMMI score of 0.88. These two agencies manage personally identifiable information (PII) such as the date of birth, social security numbers, driver's license, and identification card on a daily basis. In addition, five agencies (Behavioral Health and Developmental Services, Taxation, Aging and Rehab Services, Health, and Conservation & Recreation) also have a great responsibility to manage PII and hence at the relatively higher level of data maturity (Managed level). These agencies with the higher maturity scores could very well serve as leaders in process improvement initiatives across the enterprise.

Among the DMM categories, our analysis indicated that Platform & Architecture was at the highest level of maturity. This may follow from agencies being required to have a higher emphasis on records management even if they did not necessarily achieve a high level of data management maturity. In particular, the three process areas most closely associated with records management (Architectural Approach, Architectural Standards, and Historical Data, Retention and Archiving) have the highest competency across all agencies in the State (Figure 3). Additionally, Metadata Management has a similar score (mode = 1.75). In functionality, Metadata Management is closely associated with records management.

With regard to the Supporting Processes category, our analysis indicated that for most agencies it is at a matching level of maturity as other DMM categories. This suggests that agencies undertaking data management maturity improvements have also invested in improving Supporting Processes. However, establishing this association requires additional investigation.

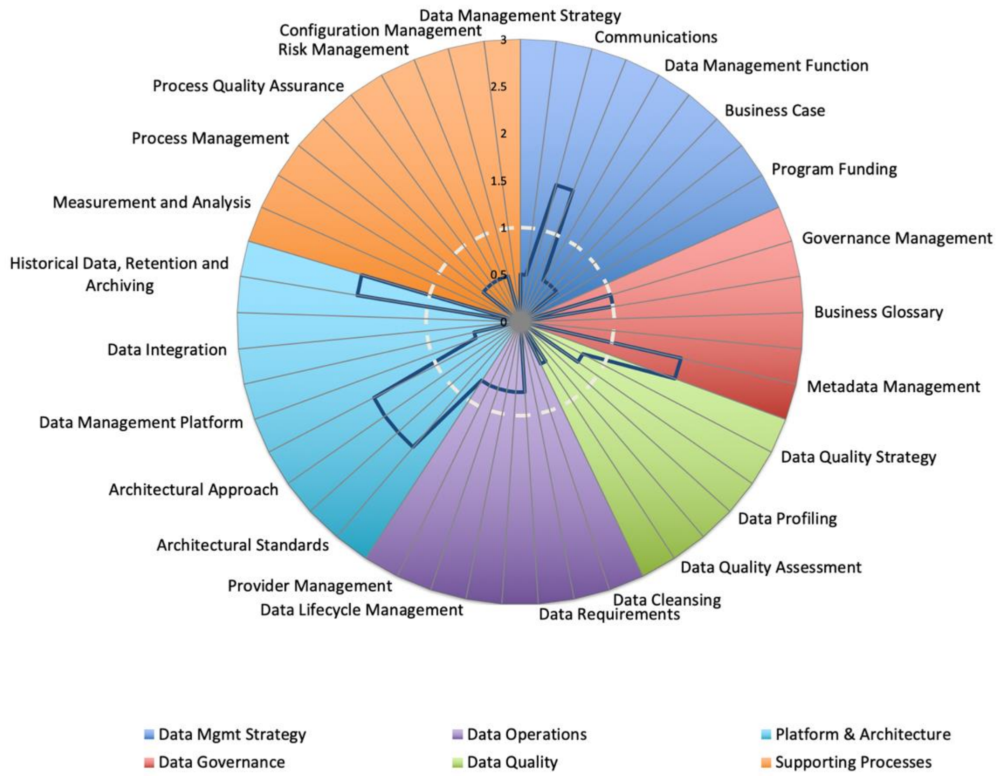


Figure 3. Overview of process area scores for all agencies

Note: Architectural approach, Architectural standards, and Historical data, retention and archiving (from Platform & Architecture), and Metadata Management (from Data Governance) have mode of 1.75. The dark blue line represents the assessment process area scores of all agencies. The white dashed line is the data management maturity Level 1.

5.2. Intra-agency assessment

Our intra-agency analysis first evaluated consistency among the DMM categories. The analysis indicated that, for most agencies that advanced beyond the lowest maturity level (i.e., *Performed*, level 1), the DMM categories were not at the same level of maturity. However, a single DMM category with a lower maturity score does not restrict the agency from achieving an overall higher maturity level. For example, in the case of Aging and Rehab Services, even though Data Operations is at level 1 (*Performed*) maturity, the overall maturity score for the agency is at level 2 (Managed). Limited resources to invest in all aspects of data management might be the explanation for the variation in DMM maturity scores. A likely explanation is that program funding is uneven. In fact, Program Funding is the least mature area within the Data Management Strategy category (see Table 2). One of the biggest challenges with program funding is that it depends not only on agency leadership, but also on the priorities of the elected representatives in the State legislature.

We then assessed the level of consistency between the DMM category maturities and the Supporting Processes category maturity. Our analysis indicated that they were independent of each other. For instance, the Supporting Processes maturity for Aging and Rehab Services was at level 5 (Optimized), yet the DMM category maturities were at level 2 (Managed). In the case of Motor Vehicle, both Supporting Processes maturity and the DMM category maturities were at level 3 (Defined). However, the Supporting Processes maturity for Taxation was at level 1 (Performed), while the DMM category maturities were at level 2 (Managed). It is to be noted that the DMM reference model framework (DMM, 2014) emphasizes the importance of Supporting Processes for DMM. However, our study suggests that Supporting Processes are not as closely associated with the adoption, execution, and improvement of data management maturity processes.

5.3. Cross-case assessment

Finally, the cross-case analysis demonstrates an approach that allows enterprises to benchmark entities based on an individual attribute such as budget or PII. Enterprises can use a similar approach to compare entities that vary in their levels of individual attributes. A benchmarking approach of this nature can be used to quantitatively assess and set priorities for improvements in efficiency and effectiveness of data management processes.

6. Implications

6.1. Implications to practice

Our research provides a frame of reference to assess and improve data management at government agencies. In the State where the study was conducted, the Governor issued an Executive Directive (ED7) in 2017 called “Leveraging the Use of Shared Data and Analytics” (VITA, 2017) to promote greater utility and accessibility of data assets maintained by the State agencies. ED7 lays out four strategic objectives linked to agency data sharing, governance, and analytics, namely, enhancing government transparency, streamlining business processes, increasing operational efficiency and effectiveness, and minimizing duplication and overlap of current and future systems development (VITA, 2017). Specifically, recommendation 2.3 of ED7 is to perform ongoing Data Management Maturity (DMM) assessments for agencies across domains of the State government. Our study provides valuable insights to State agencies to continue growth along targeted maturity curve for data management.

The DMM reference model framework does not provide a mechanism to aid the comparison of multiple entities within an enterprise. In our study, we developed the DMMI to provide a relative ranking of entities across all DMM categories. Senior level administrators can utilize this approach to aid in the comparison of business functions. The DMMI metric would enable management to identify entities that are in greater adherence to data maturity best practices. Such entities can likely serve as role models for the rest of the enterprise. The DMMI may serve as an aid to define benchmarks and drive investment decisions to improve data maturity.

The operating environment and the mission of each agency within the State are different. The DMM reference model framework was also new to the State. However, the reference model provided a common language to assess data management maturity at each agency. Overall, our study indicated that most agencies assessed were at Level 1 (Performed) and would benefit from taking steps to achieve Level 2 (Managed) and Level 3 (Defined) maturity. Achieving level 3 might sound simple, but the path towards that goal can be quite complex. We recommend an incremental approach for prioritizing resources that takes into consideration DMMI scores of each agency in conjunction with their stated missions.

Our research highlighted numerous challenges in conducting maturity assessments at State agencies. We found that many agencies do not explicitly differentiate the various process areas related to data management. Even among agencies that have initiated data management processes, lack of differentiation and broad scope of data management lead to practices that are vague and vastly abstract. A preferred approach would be for the agencies to adopt a stage model that methodologically and systematically defines the process areas and objective functions. In this regard, a stage model or a reference framework such as the DMM model provides sound footing. As discussed previously and as we show in our study, a stage model offers benefits of individually monitoring process areas, empirically evaluating current performance, setting benchmarks for longitudinally measuring progress, and developing prescriptive strategies for improvements based on pre-established performance standards.

Our interactions with the various agencies highlighted another crucial issue. Many managers viewed process maturity and data maturity as two different domains within the same organization. A likely reason is that traditional maturity models such as the CMMI were viewed as tools to assess and improve the effectiveness of process capabilities. Yet all agencies we interviewed indicated effective data management maturity as a goal to attain. Therefore, we caution managers and decision makers against a dual domain viewpoint. Our recommendation would be to combine the two domains when planning data management and improvement initiatives.

6.2. Implications for research

Solli-Sæther and Gottschalk (2010) pointed out three challenges related to developing and testing stage models - the large extent of conceptual research, the lack of empirical assessment of stages, and the practical nonexistence of linear sequences of stages in organizational life. Our research addresses the first two challenges by conducting a large-scale empirical study involving multiple agencies using an established stage model. A longitudinal study that investigates the evolution of organizational growth and its potential mis-alignment with the linear path of stage models would be required to examine the third concern.

Bannister and Connolly (2015) highlighted the difficulties associated with the use of stage models in e-government research. Key among them are: complicated to apply in government setting, limited exploration of transformational change, and the lack of depth in conceptualization. Our research addressed these difficulties. First, we used a stage model to conduct a multi-level assessment of 15 government agencies of diverse sizes, missions, and varying levels of maturity. Following J. Lee (2010) our study provides a frame of reference that can be used by researchers to evaluate stage models at government agencies. Second, our research explored how critical data assets are managed by government agencies and how well they are positioned to support organizational business goals. We proposed DMMI as a mechanism for standardization of data maturity scores and benchmarking data management against which future improvements can be assessed. Decision makers can apply DMMI to develop a roadmap for future data maturity initiatives and investments. Our research thus sets a solid foundation to further evaluate and understand transformational changes in government agencies. Third, unlike Layne et al. (2001), who emphasized a conceptualization around citizen services, we call for a comprehensive reconceptualization centered around data management maturity.

Such a reconceptualization would enable internal and external transformation. Internally, it would cushion prevailing knowledge boundaries that constrain the integration of scattered systems across agencies (Chun et al., 2010). Externally, it would assist in better organization of information, technology, and government processes for citizens' convenience.

Our research reaffirms that IS researchers are uniquely positioned to undertake theoretically founded and empirically validated studies that investigate the role of stage models in organizational growth. However, our experience showed that evaluating data maturity of a public organization can be demanding, difficult to co-ordinate, and require commitment from public officials. Yet, valuable insights of 'substantive significance' (Wacker, 1998) can emerge from the exercise that translate into recommendations for improving practice. Comprehensive knowledge of the domain, familiarity with data management concepts, and an astute scientific acumen are invaluable for researchers undertaking data maturity assessment studies. Nevertheless, we encourage future investigations to develop more and better theories (Bannister & Connolly, 2015) and to extend the body of knowledge that would benefit both practitioners and researchers.

7. Limitations and future research

Stage models are not without limitations. A notable deficiency of the CMMI DMM model is that it was not developed for comparative assessment of multiple business functions within an organization (or, in our case multiple agencies within the State). As a result, our data maturity scores do not account for variations in size, budget, and scale of data across the agencies. An astute observer may therefore call into question the suitability of the DMM model for inter-agency assessments. For example, every State would have an agency that controls vehicles and operator licenses. Because of the sheer volume and complexity of data, legal, and revenue implications, it would seem natural that the data maturity at this agency is higher than many other agencies that are less data intense. Similarly, an agency devoted to transportation and an agency devoted to health would have different data management priorities resulting in different maturity scores. Thus, an equitable comparison of data management practices across multiple business functions within an organization will be challenging without adjusting for the differences in scope of the data management initiatives, stated mission of the business unit, and complexity of the data inherent to the domain. In our research, we proposed the DMMI as a means to address this issue and to benchmark performance by standardizing for the differences. Even then, without sufficient consideration of the above-mentioned organizational factors, an inter-agency maturity assessment exercise cannot be considered holistic.

Another noteworthy limitation of any stage model is its failure to incorporate the rate of change of the organizational domain in process maturity scoring. The rate of change could vary noticeably in a rapidly evolving environment in contrast to a slowly changing environment. Maintaining data management processes is easier in the latter setting than the former. Our evaluation measured process maturity at a fixed point of time (i.e., at the time of the study) and did not gauge the rate of change of the agency’s organizational, technological, and data environment. A longitudinal reassessment would provide better insights as to how the rate of change of the organizational domain impacts data management maturity.

The DMM assessment was conducted only at one State in the United States. Comparative assessment in other States and other countries would be insightful to test the descriptive and predictive quality of stage models. Nevertheless, by providing an assessment of data maturity across diverse public agencies our study lays foundation for future theory development in related areas. Similarly, conducting assessments and comparing results using other stage models such as the Open Government Maturity Model (G. Lee & Kwak, 2012) or the DataFlux Data Governance Maturity Model (NASCIO, 2009) would be beneficial. However, our experience suggests that conducting a multi-agency assessment using multiple models would be a daunting endeavor. Nevertheless, we encourage researchers to undertake research initiatives using other methods so that a comparative analysis may be possible in the future.

Lastly, our study did not assess whether higher data process maturity leads to improvement in managerial decision making and organizational decision support. This presents a ripe area for future research.

8. Conclusion

Information and services provided by the government are increasingly accessible, transparent, and accountable to citizens, other public agencies, and external partners. Increasingly, government agencies are placing importance on the need for ongoing training and systematic assessment of data management capabilities, data governance, and data sharing practices (VITA, 2017). For example, the United States Federal Government recently passed a law to support evidence-based policy making based on open government data assets (US Gov, 2018). The law requires all government agencies to develop and maintain a comprehensive data inventory for all data assets created by or collected by the agency. Furthermore, it seeks to establish government-wide best practices for the use, protection, dissemination, and generation of data and for promoting data sharing agreements among agencies.

To fully leverage data assets, a vital goal for government agencies should be to establish higher level of data management maturity. Our study demonstrated how an enterprise-wide systematic assessment of data management maturity can be conducted using a stage model. Based on a multi-level analysis (inter-agency, intra-agency, and cross-case analysis) using the DMM reference model framework, our study draws insights on the data management capabilities at 15 State agencies. To aid the comparative assessment of multiple independent agencies, we developed and tested the DMMI. To the best of our knowledge, a multi-agency assessment of this scale has not been conducted before. Our research addresses many of the challenges associated with the use of stage models in e-government research (Bannister & Connolly, 2015; Solli-Sæther & Gottschalk, 2010). The study thus makes valuable contributions to the knowledge base. Lastly, the study provides a frame of reference and a roadmap for practitioners and decision makers involved in similar initiatives.

Appendix

Table A1: Data Management Maturity Levels (adapted from DMM, 2014)

Level	Description
1. Performed	• Processes performed ad hoc, primarily at the project level. • Processes not applied across business areas. • Primarily reactive (e.g., data quality process emphasizing repair over prevention.) • Foundational improvements may exist, but not yet extended within organization or maintained.
2. Managed	• Processes planned and executed in accordance with policies. • Processes monitored, controlled, and evaluated for adherence to policies. • Availability of skilled employees and adequate resources to produce controlled outputs. • Stakeholder engagement.
3. Defined	• Set of standards employed and consistently followed. • Processes to meet specific needs tailored from the set of standard process based on organizational guidelines.
4. Measured	• Process metrics defined and used for data management, that include: ▪ Management of variance. ▪ Analysis and prediction using statistical and quantitative techniques. • Process performance managed across the lifespan of the process.
5. Optimized	• Process performance optimized through Level 4 (Performed) analysis for identification of improvement opportunities. • Best practices shared with peers and industry.

References:

Andersen, K. V., & Henriksen, H. Z. (2006). E-government maturity models: Extension of the Layne and Lee model. Government information quarterly, 23(2), 236-248.

Anderson, C. R., & Zeithaml, C. P. (1984). Stage of the product life cycle, business strategy, and business performance. Academy of Management journal, 27(1), 5-24.

Bannister, F., & Connolly, R. (2015). The great theory hunt: Does e-government really have a problem? Government information quarterly, 32(1), 1-11.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing maturity models for IT management. Business & Information Systems Engineering, 1(3), 213-222.

Benlian, A., & Hess, T. (2011). Opportunities and risks of software-as-a-service: Findings from a survey of IT executives. Decision Support Systems, 52(1), 232-246.

Bradley, R. V., Pratt, R. M., Byrd, T. A., Outlay, C. N., & Wynn Jr, D. E. (2012). Enterprise architecture, IT effectiveness and the mediating role of IT alignment in US hospitals. Information Systems Journal, 22(2), 97-127.

Brookes, N., Butler, M., Dey, P., & Clark, R. (2014). The use of maturity models in improving project management performance: An empirical investigation. International Journal of Managing Projects in Business, 7(2), 231-246.

Carmel, E., & Agarwal, R. (2006). The maturation of offshore sourcing of information technology work. In Information systems outsourcing (pp. 631-650): Springer.

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347.

Chun, S., Shulman, S., Sandoval, R., & Hovy, E. (2010). Government 2.0: Making connections between citizens, data and government. Information Polity, 15(1, 2), 1-9.

CMMI. (2002). Capability Maturity Model® Integration (CMMI), Version 1.1--Continuous Representation. (CMMI Product Team).

CMMI. (2016). Data Management Maturity Model. Retrieved from http://cmmiinstitute.com/data-management-maturity

Comuzzi, M., & Patel, A. (2016). How organisations leverage big data: A maturity model. Industrial Management & Data Systems, 116(8), 1468-1492.

Dennis, A. L. (2018). Case Study: State of Arizona Implements Model for State-Wide Data Governance. Retrieved from https://www.dataversity.net/case-study-state-arizona-implements-model-state-wide-data-governance/#

DMM. (2014). Data Management Capability Assessment Model. Retrieved from http://cmmiinstitute.com/data-management-maturity

Galliers, R. D., & Sutherland, A. (1991). Information systems management and strategy formulation: the ‘stages of growth’model revisited. Information Systems Journal, 1(2), 89-114.

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. Decision Support Systems.

Gottschalk, P. (2009). Maturity levels for interoperability in digital government. Government information quarterly, 26(1), 75-81.

H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2017, Pub. L. No. 115th Congress of the United States (2017-2018) (2018).

Hiller, J. S., & Bélanger, F. (2001). Privacy strategies for electronic government. E-government, 200, 162-198.
Iversen, J., Nielsen, P. A., & Norbjerg, J. (1999). Situated assessment of problems in software development. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 30(2), 66-81.
Janowski, T. (2015). Digital government evolution: From transformation to contextualization. Government information quarterly, 32(3), 221-236.
Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. Journal of Business Research, 70, 338-345.
Khan, S., & VanWynsberghe, R. (2008). Cultivating the under-mined: Cross-case analysis as knowledge mobilization. Paper presented at the Forum Qualitative Sozialforschung/Forum: Qualitative Social Research.
Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3), 78-85.
King, J. L., & Kraemer, K. L. (1984). Evolution and organizational information systems: an assessment of Nolan's stage model. Communications of the ACM, 27(5), 466-475.
Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. Government information quarterly, 18(2), 122-136.
Lee, G., & Kwak, Y. H. (2012). An open government maturity model for social media-based public engagement. Government information quarterly, 29(4), 492-503.
Lee, J. (2010). 10 year retrospect on stage models of e-Government: A qualitative meta-synthesis. Government information quarterly, 27(3), 220-230.
Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. Government information quarterly, 29(4), 446-454.
Maheshwari, D., & Janssen, M. (2013). Measurement and benchmarking foundations: Providing support to organizations in their development and growth using dashboards. Government information quarterly, 30, S83-S93.
Mathison, S. (2005). Encyclopedia of evaluation. Thousand Oaks, California: Sage Publications.
NASCIO. (2009). Data Governance Part II: Maturity Models - A Path to Progress. Retrieved from Lexington, KY:
Nolan, R. L. (1973). Managing the computer resource: a stage hypothesis. Communications of the ACM, 16(7), 399-405.
Nolan, R. L. (1979). Managing Crises of Data Processing. Harvard business review, 3(4).
Norris, D. F., & Lloyd, B. A. (2006). The scholarly literature on e-government: Characterizing a nascent field. International Journal of Electronic Government Research (IJEGR), 2(4), 40-56.
Obama, B. (2009). Memorandum for the heads of executive departments and agencies. Presidential Studies Quarterly, 39(3), 429.
Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. Information & Management, 51(5), 497-510.
Pardo, T. A., Nam, T., & Burke, G. B. (2012). E-government interoperability: Interaction of policy, management, and technology dimensions. Social Science Computer Review, 30(1), 7-23.

Poeppelbuss, J., Niehaves, B., Simons, A., & Becker, J. (2011). Maturity models in information systems research: Literature search and analysis. CAIS, 29(1), 1-15.
Röglinger, M., Pöppelbuß, J., & Becker, J. (2012). Maturity models in business process management. Business Process Management Journal, 18(2), 328-346.
Scott, M., & Bruce, R. (1987). Five stages of growth in small business. Long range planning, 20(3), 45-52.
Solli-Sæther, H., & Gottschalk, P. (2010). The modeling process for stage models. Journal of Organizational Computing and Electronic Commerce, 20(3), 279-293.
Swinarski, M., Parente, D. H., & Kishore, R. (2012). Do small IT firms benefit from higher process capability? Communications of the ACM, 55(7), 129-134.
Vezzetti, E., Violante, M. G., & Marcolin, F. (2014). A benchmarking framework for product lifecycle management (PLM) maturity models. The International Journal of Advanced Manufacturing Technology, 71(5-8), 899-918.
VITA. (2017). Executive Directive 7: Leveraging the Use of Shared Data and Analytics. Retrieved from https://www.vita.virginia.gov/media/vitavirginiagov/it-governance/pdf/ExecutiveDirective7.pdf
Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. Journal of operations management, 16(4), 361-385.
Yin, R. K. (2009). Case study research: Design and methods. Thousand Oaks, California: Sage publications.