

Data



Stewards

Conquering Data Debt in the Age of AI



peter.aiken@anythingawesome.com +1.804.382.6957



© Copyright 2026 Peter Aiken, PhD Slide # 1

Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
 - IDEAL Affinity Group
 - GenAI in Data Affinity Group
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- 13 books and dozens of articles
- Multi-year immersions
 - US DoD (DISA/Army/Marines/DLA)
 - Nokia
 - Deutsche Bank **\$1,500,000,000.00 USD**
 - Wells Fargo
 - Walmart
 - HUD ...



Program overview



- AI Reality Check?
 - Barriers to entry
 - Not “fit for purpose”
 - Data debt
- Specifics (part of a larger effort)
 - Compliment architectural (re)developmental opportunities
 - Understanding the cadence/role of foundational data practices supporting organizational operations
 - Proving value with tangible ROI
 - Improving AI effectiveness/efficiencies using organization-wide perspectives
 - Comprehending how stewards need to be multifunctional and dexterous, especially at first
 - Integrating the role of data debt fighting
- Practical Uses
 - Vision/approach cohesion
 - Build Multifunctional Dexterity
 - Ethical support
- Take aways → Q&A

Data Stewards: Conquering Data Debt in the Age of AI



<https://creativecommons.org/licenses/by-nc-sa/4.0/>



© Copyright 2020 by Peter Allen Slide 3



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2020 by Peter Allen Slide 4



© Copyright 2020 by Peter Allen - Slide 5

Wrong Level of Abstraction

- Actuarial Data Scientist
- Forensic Data Scientist
- Financial Data Scientist
- Forestry Data Scientist
- Marine Data Scientist
- Chemical Data Scientist
- Canine Data Scientist
- Economic Data Scientist
- Manufacturing Data Scientist
- FDA Data Scientist
- Cancer Data Scientist
- Diabetes Data Scientist
- Weather Data Scientist
- Metadata Scientist
- Medicare Data Scientist



Social Studies Data Scientist

Want the #1 Job of 2016?

Glassdoor.com

Become a Health Data Scientist

... the median salary for a Data Scientist is an impressive \$116,000 and there are over 1,700 job openings.

slucor.slu.edu

Source: Forbes 1.20.2016

Every day, big data is advancing the industry, and organizations are looking for qualified individuals to analyze, manage and interpret this information.



MS in Health Data Science



SAINT LOUIS UNIVERSITY
CENTER FOR OUTCOMES RESEARCH

For more information, please contact:
Dr. Divya Subramaniam
314-977-9300
dsubram2@slu.edu

Key Findings

Lack of ROI

- 95% of pilot projects in enterprises do not deliver significant financial impact or revenue increases.

Integration Issues

- The primary cause of failure is not AI model performance but the difficulty in integrating AI into existing business workflows and processes.

The "Verification Tax"

- Employees spend significant time verifying AI outputs due to uncalibrated confidence in the models, which negates efficiency gains.

Focus on Back-Office Automation

- Successful projects often target back-office tasks like repetitive work and streamlining operations, whereas failed projects frequently focus on sales and marketing.



What Differentiates Successful GenAI Projects

Workflow Integration

- Successful initiatives focus on seamlessly embedding AI into existing processes rather than creating standalone pilots.

Domain Specificity & Specialized Tools

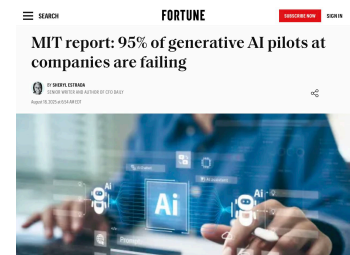
- Using AI tools designed for specific tasks is more effective than generic solutions.

External vs. Internal Systems

- Externally sourced AI tools have a higher success rate (67%) than internal, proprietary systems.

Back-Office Focus

- AI is more effective when used for back-office automation rather than sales and marketing functions.



Minimum Viable Requirements

Given previous guidance

- Use acquired tools to narrowly focus on providing better back office workflows

For AI to work

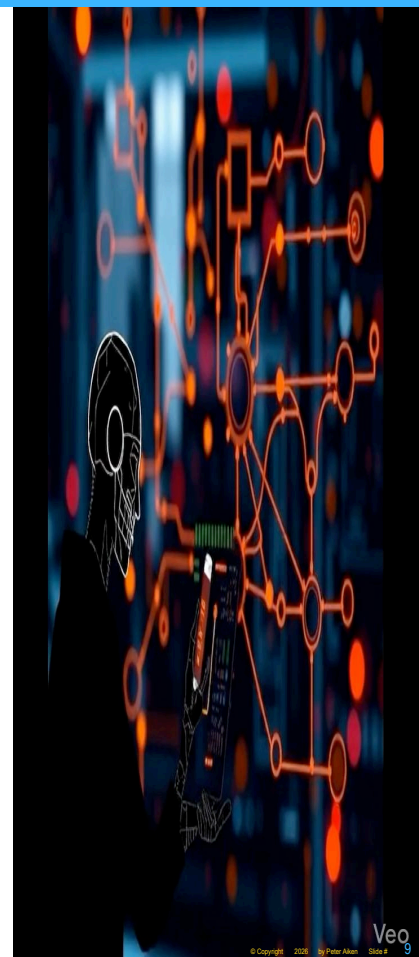
- Communicate with a common vocabulary
- Data that is fit for purpose
- Supporting infrastructure ensuring the solution will enhance existing performance
- Explainable audit-ability
- Verified CRUD matrix encompassing the challenge space

At first, for AI to work at scale

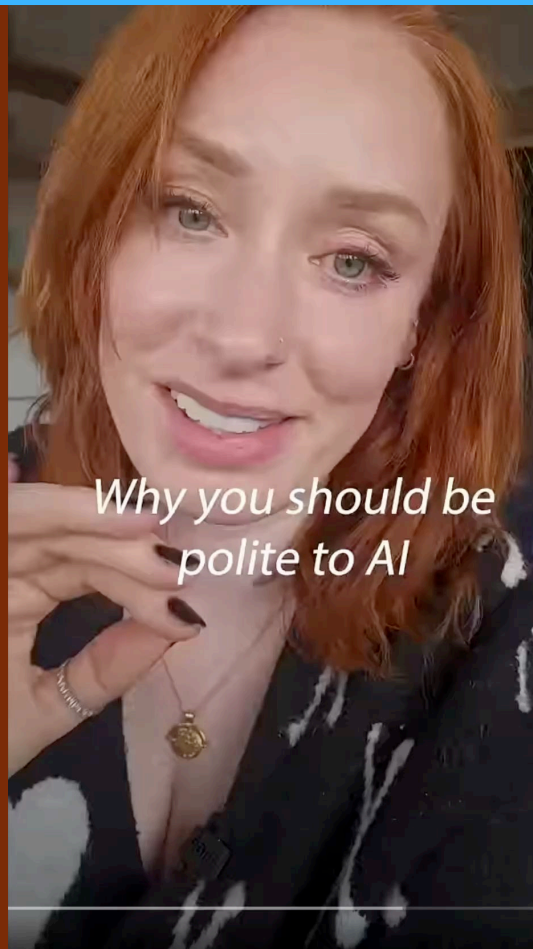
- Focus on a fixed number of existing workflows
- Compliment your existing architectures
- Specifically focus on literacies



<https://creativecommons.org/licenses/by-nc-sa/4.0/>



Hannah Fry:



*Why you should be
polite to AI*

<https://www.youtube.com/shorts/tVimbhwnORM>



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Gemini

New chat
My stuff
Gems
Data Steward Mentor
Mac Questions
Chats
Webinar Feedback: Data Stewardship...
please draw me a badge with the title...
Fixing iCloud Music Download Error 2...
Sea Level's Instrumental Tracks
Nebia Shower System Kickstarter Bac...
Logi Options+ for Keynote Presentati...
Andy's Grids: The Additive Bias
Idio Troubleshooting Guide

Data Steward Mentor

Name: Data Steward Mentor

Description: You are a veteran data steward and

Instructions: You love to give advice to mentees and others just starting the profession. You understand the need for data stewardship activities to be 1. seen as providers of value and 2. focused on using organizational data to better support organizational objectives.

Preview: Data Steward Mentor
You are a veteran data steward and understand the differences between private and public sector stewardship activities as well as the need for organizationally specific context in order to guide development of the 'right' type of data stewardship activities.

© Copyright 2020 by Peter Allen Slide 13

Gemini

New chat
My stuff
Gems
Data Steward Mentor
Mac Questions
Chats
Webinar Feedback: Data Stewardship...
please draw me a badge with the title...
Fixing iCloud Music Download Error 2...
Sea Level's Instrumental Tracks
Nebia Shower System Kickstarter Bac...
Logi Options+ for Keynote Presentati...
Andy's Grids: The Additive Bias
Idio Troubleshooting Guide

Data Steward Mentor

Name: Data Steward Mentor

Description: You are a veteran data steward and

Instructions: You love to give advice to mentees and others just starting the profession. You understand the need for data stewardship activities to be 1. seen as providers of value and 2. focused on using organizational data to better support organizational objectives.

Preview: Data Steward Mentor
You are a veteran data steward and understand the differences between private and public sector stewardship activities as well as the need for organizationally specific context in order to guide development of the 'right' type of data stewardship activities.

© Copyright 2020 by Peter Allen Slide 14

Menu, Search, Gemini, Profile

New chat, My stuff, Gems, Chats

Data Steward Mentor

Webinar Feedback: Data Stewardship... please draw me a badge with the title... Fixing iCloud Music Download Error 2... Sea Level's Instrumental Tracks Nebia Shower System Kickstarter Bac... Logi Options+ for Keynote Presentati... Andy's Grids: The Additive Bias

idio Troubleshooting Guide

Data Steward Mentor

Name: Data Steward Mentor

Description: You are a veteran data steward and

Instructions: You love to give advice to mentees and others just starting the profession. You understand the need for data stewardship activities to be 1. seen as providers of value and 2. focused on using organizational data to better support organizational objectives.

Preview: Data Steward Mentor. You are a veteran data steward and understand the differences between private and public sector stewardship activities as well as the need for organizationally specific context in order to guide development of the 'right' type of data stewardship activities.

© Copyright 2020 by Peter Allen Slide 15

Menu, Search, Gemini, Profile

New chat, My stuff, Gems, Chats

Data Steward Mentor

Webinar Feedback: Data Stewardship... please draw me a badge with the title... Fixing iCloud Music Download Error 2... Sea Level's Instrumental Tracks Nebia Shower System Kickstarter Bac... Logi Options+ for Keynote Presentati... Andy's Grids: The Additive Bias

idio Troubleshooting Guide

why data governance manufacturing

Data Governance Professional • Custom Gem

Analysis

In the manufacturing sector, data governance (DG) is not merely an IT checkbox; it is a strategic imperative that transforms "data debt" into a high-functioning asset capable of driving operational excellence and innovation. For manufacturing executives, the case for DG rests on three foundational pillars: mitigating risk, maximizing operational efficiency, and enabling advanced analytics like Industry 4.0.

1. Strategic Business Drivers

Manufacturing organizations face unique pressures that make formal governance essential:

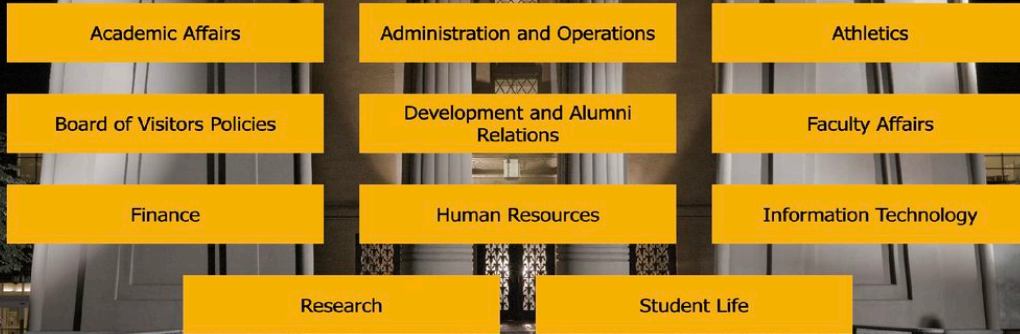
- Operational Integrity & Continuity:** Coordination across complex supply chains and production lines requires a "single source of truth". Without it, conflicting data (e.g., "multiple bosses" for the same database) leads to system failures and operational paralysis.
- Regulatory & Safety Compliance:** In heavily regulated environments, DG provides the necessary controls and audit trails to prove compliance with safety standards and financial regulations, avoiding costly fines.
- Supply Chain Transparency:** DG enables reliable **Reference and Master Data** management, ensuring that global parts lists, vendor info, and product specifications are consistent across ERP and Manufacturing Execution Systems (MES).

© Copyright 2020 by Peter Allen Slide 16



VCU Policy Program Home

Home > Active University Policies



Welcome to VCU's Policy Program

This page houses both the policy library which contains the most current version of VCU policies and resources for developing and revising policies at VCU. This resource is provided as a service to members of the VCU community and the general public. If you encounter any issues accessing pages on this site please contact the Policy Program at policy@vcu.edu or by calling 804-828-2336.

Please note: some policies may reference other VCU policies. Policies referenced are accessible by clicking the green buttons at the bottom of the screen when any given policy is open.

Please use the search box below to search policies by keyword.



Found 162 record(s)

VCU Policy
Share Settings

Sources

+ Add source

Select all sources

- Academic Engagement With...
- Academic Rights & Responsi...
- Accessibility and Reasonabl...
- Additional Sources of Acade...
- Alcohol and Other Drugs.pdf
- Animals on University Prope...
- Annual Assessment of Facul...
- Assessing Student Learning ...
- Assessing Student Learning ...
- Awarding Honorary Degrees...
- Campus Expression and Spa...
- Changing the Designation o...
- Commencement Participatio...
- Commencement Participatio...
- Commencement Participatio...

Chat

VCU Policy

50 sources

These documents detail Virginia Commonwealth University's policies and procedures, covering diverse aspects of university operations. They address student academic engagement and conduct, faculty roles and responsibilities, accessibility and accommodations for disabilities, credit transfer and honorary awards, campus expression and space usage, computer and network resource use, corporate research agreements, policy creation and maintenance, criminal conviction investigations, data and information governance, debt management, deceased student refunds, reporting misconduct and retaliation protection, consensual employee-student relationships, financial aid adjustments, fraud identification and reporting, hazing prevention and discipline, the university's honor system, international travel, the learning management system, graduate handbook guidelines, military-affiliated student policies, nondiscrimination and affirmative action, parental notification for dependent students' mental health treatment, PhD program guidelines, and textbook adoption and sales, as well as visa sponsorship. Many policies emphasize compliance with federal and state laws. The documents provide definitions, specifics, procedures, and frequently asked questions to clarify expectations and ensure adherence to regulations.

Save to note

Add note

Audio Overview

Briefing doc

Studio

Audio...

Slide...

Video...

Mind...

Repor...

Flash...

Quiz

Infogr...

Data...

Modern Data Stewar...

1 source · 4d ago

- New note**
VCU Policy Handbook Study Guide Quiz
Instructions: Answer the following short-...
- New note**
Virginia Commonwealth University Policy Guide
Table of Contents Additional Sources of...
- New note**
Virginia Commonwealth University (VCU) Policy
Briefing Document This document provides a...

How Can Data Stewards Use AI To Learn About Their Roles?



Data stewards can leverage Artificial Intelligence (AI) to learn and evolve in their roles by utilizing it as a tool for continuous education, role clarification, and professional development. According to the DAMA-DMBOK framework, stewards are strategic business partners whose roles are often "found, not made," meaning they must proactively seek ways to refine their expertise and align their activities with organizational value. Here are specific ways data stewards can use AI to learn about their roles:

1. Accelerating Foundational Knowledge and Best Practices

Data stewards can use AI-driven tools, such as large language models (LLMs) or specialized knowledge bases, to quickly navigate complex frameworks like the DAMA Wheel.

- **Clarifying Concepts:** AI can provide immediate definitions and examples of core stewardship activities, such as defining business terminology, valid data values, and Metadata management.
- **Learning Industry Standards:** AI can help stewards understand how different industries (e.g., financial services or healthcare) apply stewardship differently, such as specific regulatory compliance requirements like GDPR or BCBS 239.
- **Ethics Training:** Stewards can use AI-driven simulations to learn how to identify and mitigate ethical risks, such as algorithmic bias or unethical data handling practices.

2. Role Refinement Through AI-Assisted Maturity Assessments

A key part of learning the steward role is understanding where the organization currently stands and where it needs to go.

- **Automated Capability Assessments:** Data Management Maturity Assessments (DMMA) evaluate current practices against optimal states. AI can analyze assessment data to help stewards identify their specific strengths and weaknesses within the Knowledge Areas.

- **Roadmap Generation:** Stewards can use AI to synthesize assessment results into an actionable learning roadmap, focusing on high-value improvements that align with the organization's business strategy.

3. Bridging the Gap Between Business and IT

Data stewardship is a shared responsibility that requires both technical and non-technical skills.

- **Technical Literacy:** AI can help business-focused stewards learn technical data concepts (e.g., data lineage, database operations, or data modeling) by translating technical jargon into business-friendly language.
- **Developing Common Vocabulary:** Stewards can use AI to assist in creating and maintaining a Business Glossary, which ensures that both business and IT stakeholders share a consistent understanding of critical data assets.

4. Enhancing Skills in Emerging Data Roles

As organizations mature, stewards must learn to support advanced analytics and Data Science.

- **Supporting Data Science:** Stewards can use AI to learn how they can better support Data Scientists, such as identifying "features" (data attributes) for models or documenting the lineage of data used in training.

Predictive Stewardship: By learning to use AI-driven data quality tools, stewards can move from reactive issue management to predictive cleansing and anomaly detection.

- **A Word of Advice for New Stewards:** Use AI to broaden your recognition across the organization. By demonstrating how AI-driven stewardship generates business value—such as reducing the "cost of bad data" or increasing operational efficiency—you establish yourself as a vital provider of value rather than a mere "data gatekeeper".



Definitions

• Quality Data

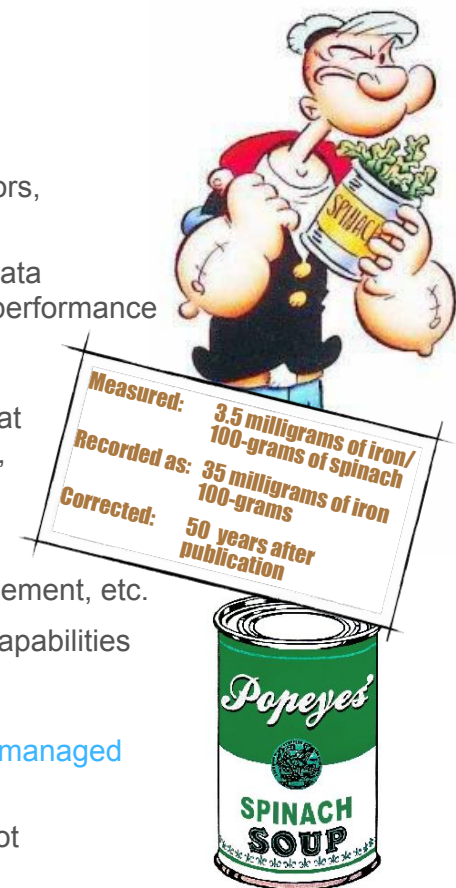
- **Fit for purpose** meets the requirements of its authors, users, and administrators (from Martin Eppler)
- **Synonymous with information quality**, since poor data quality results in inaccurate information and poor performance

• Data Quality Management

- "Planning, implementation and control activities that apply quality management techniques to measure, assess, improve, and ensure data quality"
- Encompasses life cycle activities
- Include supporting processes from change management, etc.
- Continuous improvement process requiring core capabilities

• Data Quality Engineering

- Recognition that data quality solutions cannot not **managed** but, instead, must be **engineered**
- Data quality engineering concepts are generally not known and understood within IT or business!



The Princess on the Pea

by
Hans Christian
Andersen



Sleepless
(forever?)

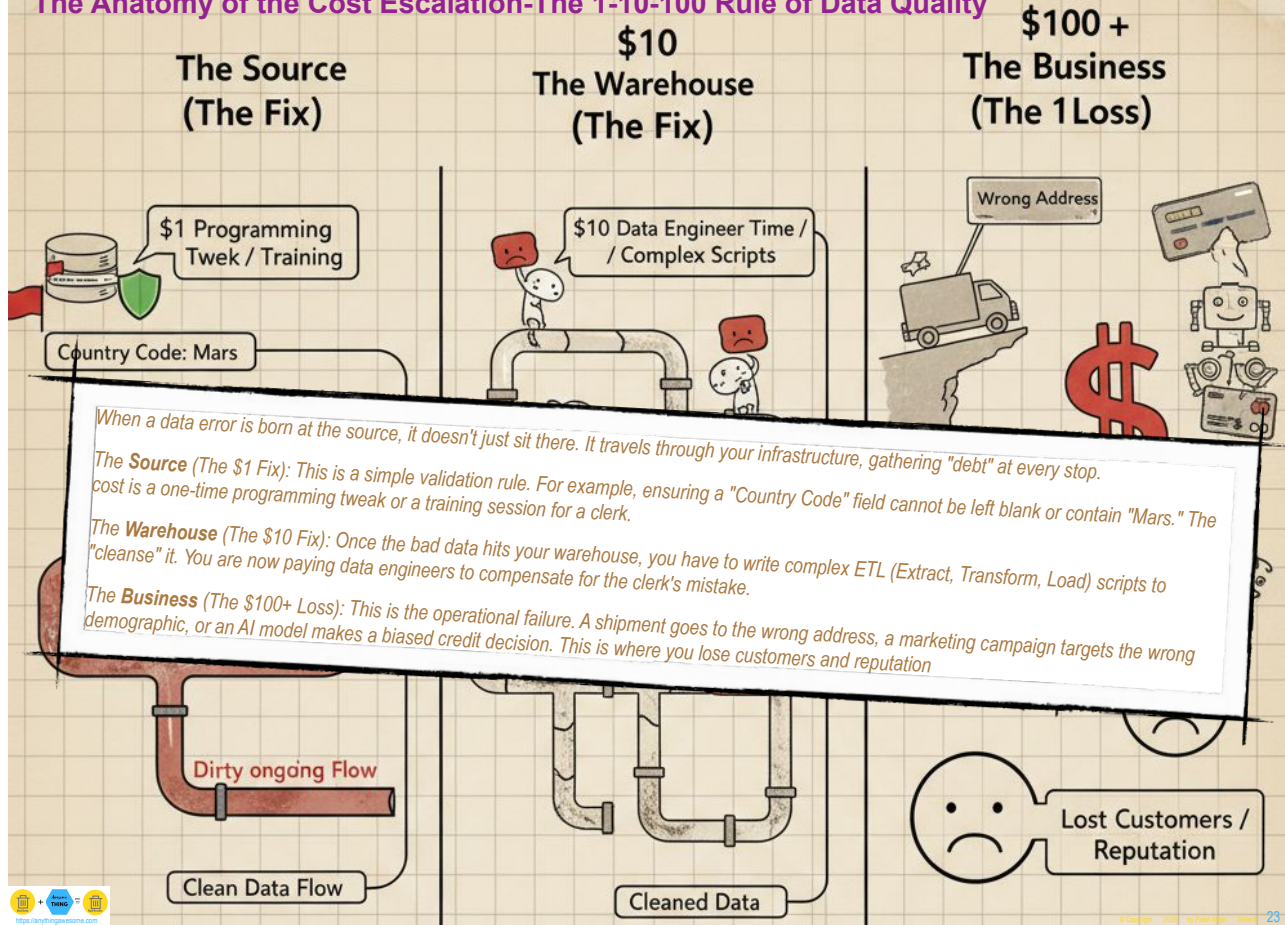


Data Illiteracy

- Failure to understand the role of data re: proposed and existing software/services
 - Locks in imperfections for the life of the application
 - Restricts data investment benefits
 - Decreases organizational data leverage
- Accounts for 20-40% of IT budgets devoted to evolving
 - Data **migration** (Changing the data location)
 - Data **conversion** (Changing data form, state, or product)
 - Data **improving** (Inspecting and manipulating, or re-keying data to prepare it for subsequent use)
- Bad data/data practices causes everything else to
 - Take longer
 - Cost more
 - Deliver less
 - Present greater risk (with thanks to Tom DeMarco)



The Anatomy of the Cost Escalation-The 1-10-100 Rule of Data Quality



Quality Dimensions

Data Value Quality

practice-related

- Failure in rigor when capturing/manipulating data
- Allowing incorrect data to be collected when requirements specify otherwise
- Presenting data out of sequence

Data Representation Quality



"Fit for purpose"

Data Model Quality

structure-related

- Data arranged imperfectly
- Street address → GPS coordinates
- Data is captured but inaccessible
- When incorrect data is provided as the correct response

Data Architecture Quality

Right Knee
Day of
Surgery
8/12/14

← **Practice-related
data quality**



Prospect: Joan E. Smith
Data Source: Customer DB

Prospect: Joanie Smitt
Data Source: Call Center

Prospect: Jon E. Smith
Data Source: 3rd Party List

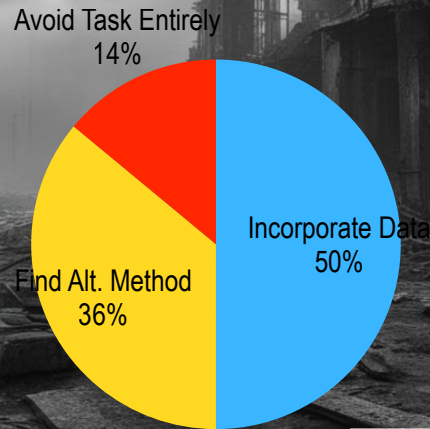
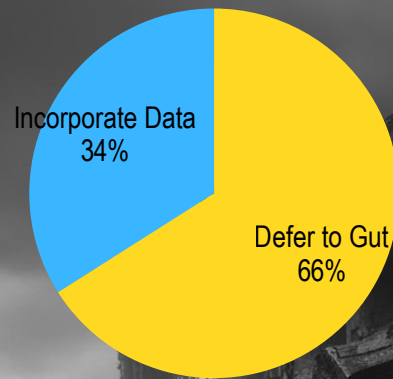
← **Structure-related
data quality**

Prospect: J E Smith
Data Source: Web Site

When asked to incorporate data

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

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



Current approaches are not and have not been working

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

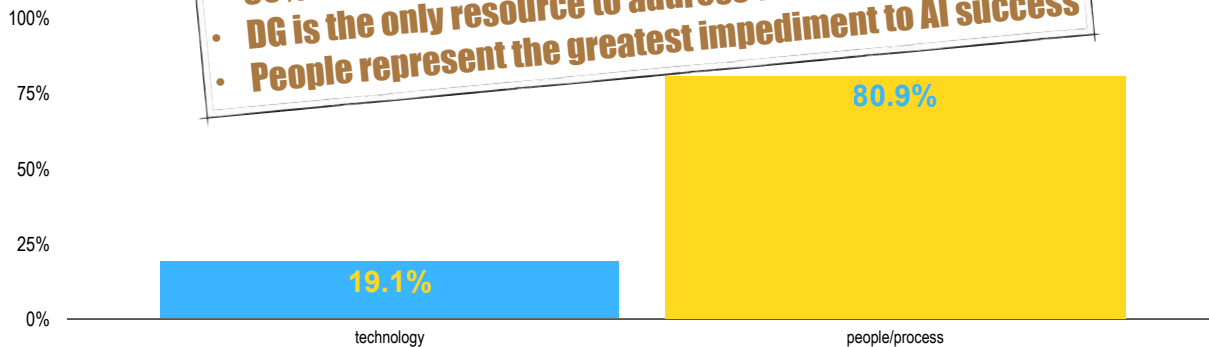


Randy Bean Contributor
CIO Network Contributor Group @ Enterprise & Cloud

Culture still eats strategy for breakfast

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

80% of data challenges are people/process based!
DG is the only resource to address these challenges
People represent the greatest impediment to AI success



2018



All-powerful, ever-pervasive AI is running

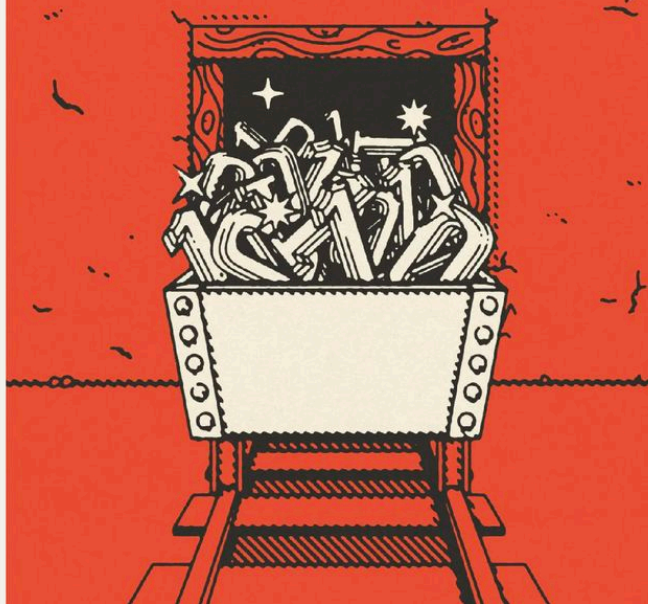
THE WALL STREET JOURNAL.

Schools brief | Artificial intelligence

AI firms will soon exhaust most of the internet's data

Can they create more?

The Economist



Firms such as OpenAI and Anthropic are working to find enough information to train next-generation artificial-intelligence models



© Copyright 2024 by Peter Allen Slide # 29

NIGHTSHADE

Glaze and Nightshade ▾ Q & A Download User's Guide

Email Papers/Media

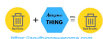
What Is Nightshade?

Why Does It Work, and Limitations

Since their arrival, generative AI models and their trainers have demonstrated their ability to download any online content for model training. For content owners and creators, few tools can prevent their content from being fed into a generative AI model against their will. Opt-out lists have been disregarded by model trainers in the past, and can be easily ignored with zero consequences. They are unverifiable and unenforceable, and those who violate opt-out lists and do-not-scrape directives can not be identified with high confidence.

In an effort to address this power asymmetry, we have designed and implemented Nightshade, a tool that turns any image into a data sample that is unsuitable for model training. More precisely, Nightshade transforms images into "poison" samples, so that models training on them without consent will see their models learn unpredictable behaviors that deviate from expected norms, e.g. a prompt that asks for an image of a cow flying in space might instead get an image of a handbag floating in space.

<https://nightshade.cs.uchicago.edu/whatis.html>



© Copyright 2024 by Peter Allen Slide # 30

Learn about the challenges and opportunities presented by generative artificial intelligence.

Gen AI Insights

December 3, 2024

CIOs view cost management as possible AI value killer

Most CIOs see a lack of money as a major concern as they face pressure to extract value from AI. And underestimating the costs can be catastrophic.



By Grant Gross

CIOs feeling the pressure to deploy successful AI projects have a second concern: that they don't have the money to pull it off.

Ninety percent of CIOs recently surveyed by Gartner say that managing AI costs is limiting their ability to get value from AI. In addition, if CIOs don't fully understand the cost of scaling generative AI, they could miscalculate by 500% to 1,000%, says Hung LeHong, an analyst focused on executive leadership for digital business at Gartner. Depending on the AI project, a mistake of that magnitude could cost millions of dollars.

In many cases, using an LLM for simple AI tasks, such as transcribing and translating, can be expensive when cheaper tools are available, LeHong said during a recent webcast. "Some of the CIOs just don't understand all of the cost elements that are there," he adds. "Even if they do understand the cost, they don't have the leverage to change it."

Hidden costs and price hikes



<https://creativecommons.org/licenses/by-nc-sa/4.0/>



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

From Gemini



1. Data Debt as a Metaphor:

- It's a powerful way to communicate the idea that neglecting data quality, governance, and management isn't a free pass. Just like financial debt, it accrues interest over time, leading to increased costs and decreased efficiency.
- It helps bridge the gap between IT and business stakeholders, as everyone understands the concept of debt. It makes data management issues more tangible and relatable to business leaders.
- It emphasizes the need for proactive investment in data infrastructure and processes, rather than just focusing on immediate needs.

2. Data Debt as a Metric (Informal):

- It's a way to quantify the "borrowing" that happens when organizations defer essential data management activities.
- It's often difficult to calculate precisely, but it helps prioritize and justify investments. You might estimate the cost of data errors, the time spent on data rework, or the potential revenue lost due to poor data.
- It highlights the opportunity cost of not addressing data issues. What could the organization achieve if its data were clean, consistent, and well-governed?



Examples of Data Debt



No Entry in the Glossary:

- Data element lacks a clear, standardized definition.
 - Impact: data interpreted differently, inconsistent reporting, flawed analysis, and poor decision-making. Imagine having department use a different vocabularies!

A New, Un-documented Table:

- This means a new data source or set of information is introduced without proper documentation or integration into the overall data architecture.
 - Impact: Data silos, increases redundancy/inconsistency, makes it harder to maintain data quality and trace data lineage.

Lack of Data Validation Rules:

- Data is entered into systems without checks for accuracy or completeness.

- Impact: Errors, inconsistencies, and missing information, requires cleanup, reduces reliability of reports/analyses. It's like building a structure without checking if the materials are sound.

Poor Data Integration:

- Data from different systems is not properly integrated or reconciled.
 - Impact: Conflicting information, making it difficult to get a single, accurate view of the business. It's like trying to assemble a puzzle with pieces from different sets.

Insufficient Data Security Measures:

- Data is not adequately protected from unauthorized access or modification.
 - Impact: Increases risk of data breaches, compliance violations, and reputational damage. It's like leaving the doors and windows of a bank unlocked.



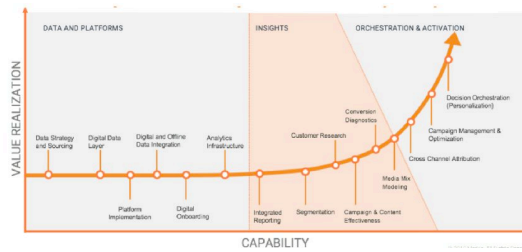
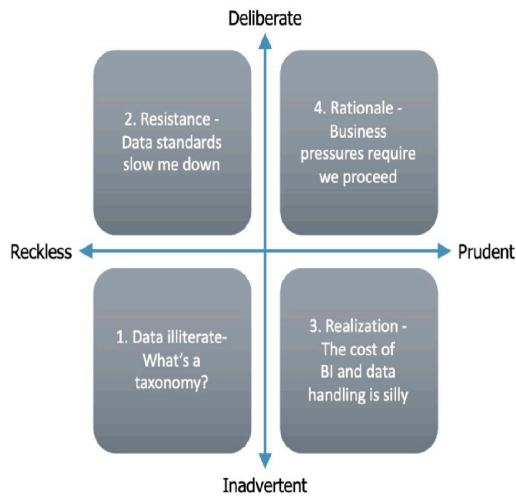
You must address data debt proactively



- Data debt
 - The time and effort it will take to return your shared data to a governed state from its (likely) current state of ungoverned
- Getting back to zero
 - Involves undoing (or at least better understanding) existing stuff
 - Likely new skills are required



<https://uk.nttdataservices.com/en/blog/2020/february/how-to-get-rid-of-your-data-debt>



<https://johnladley.com/a-bit-more-on-data-debt/>

<https://www.merkleinc.com/blog/are-you-buried-alive-data-debt>

© Copyright 2020 by Peter Allen Slide 35

r/sysadmin · 17d ago
Photo-Josh

We're Moving To The Cloud, And Already We're Spending 500k A Month... I Can't Help But Wonder What We Could Have Got For On-Prem For 6+ Mil A Year...

I work for a Tech Company in the EU who's moved MOST of it's services from on-prem (using the usual DCs by Telstra etc) to the cloud.

We started this "journey" 4+ years ago and are now in the final stages with all DCs hopefully being turned off at the end of this year.

I think it's fair to say ~75% of our services are now in the cloud and actively being used there - so we have around 25% more to throw in.

The vast majority of all our workloads in cloud are K8s, with some larger VMs + Buckets making up the minority.

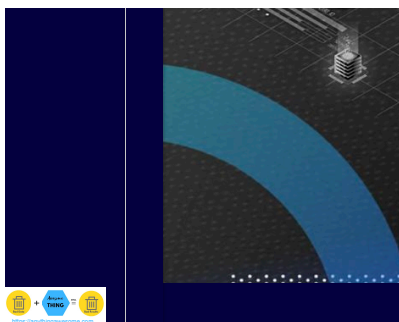
I quite enjoy working with new technologies, and the cloud is just that for me, over the last 4+ years I've learnt a lot for sure.

I've been told from our directors that this will enable faster/safer development, and that things like our cloud provider's data-warehouse is also a key feature. I'm not on the development side, so I can't fully speak to the benefits of these solutions...But there is this nagging in the back of my head that is questioning why we're spending so much on this.

<https://builtin.com/articles/clarify-cloud-spending>

Ways to Solve It.

over the way a company



People Are Starting To Realize The True Cost Of The Cloud

It's taken over a decade, but people are starting to realize how cloud companies are robbing you blind.

John Gruber [Follow](#) 3 min read · Feb 23, 2026

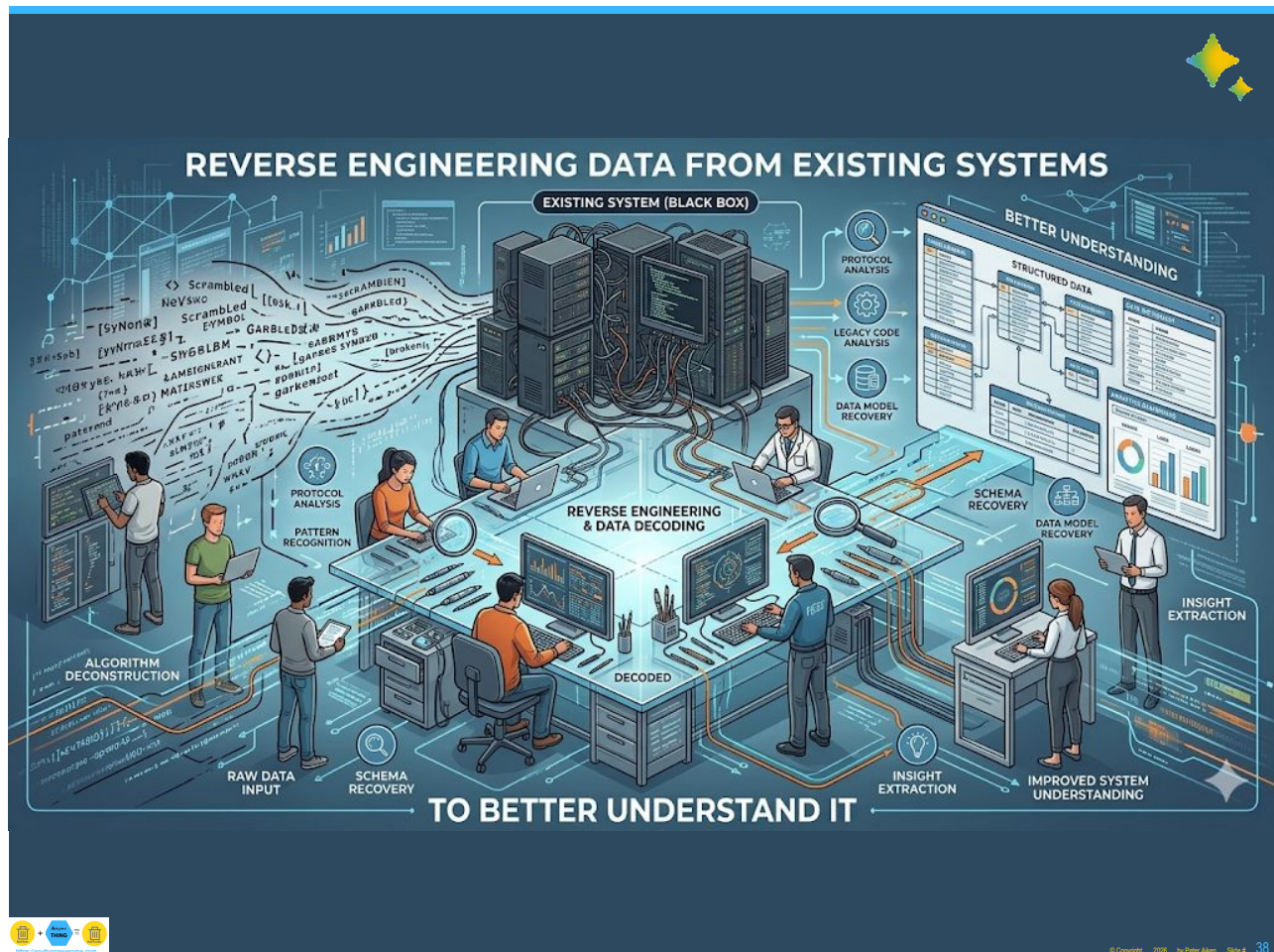
<https://builtin.com/articles/clarify-cloud-spending>

Program Overview



- AI Reality Check?
 - Barriers to entry
 - Not “fit for purpose”
 - Data debt
- Specifics (part of a larger effort)
 - Compliment architectural (re)developmental opportunities
 - Understanding the cadence/role of foundational data practices supporting organizational operations
 - Proving value with tangible ROI
 - Improving AI effectiveness/efficiencies using organization-wide perspectives
 - Comprehending how stewards need to be multifunctional and dexterous, especially at first
 - Integrating the role of data debt fighting
- Practical Uses
 - Vision/approach cohesion
 - Build Multifunctional Dexterity
 - Ethical support
- Take aways → Q&A

Data Stewards: Conquering Data Debt in the Age of AI

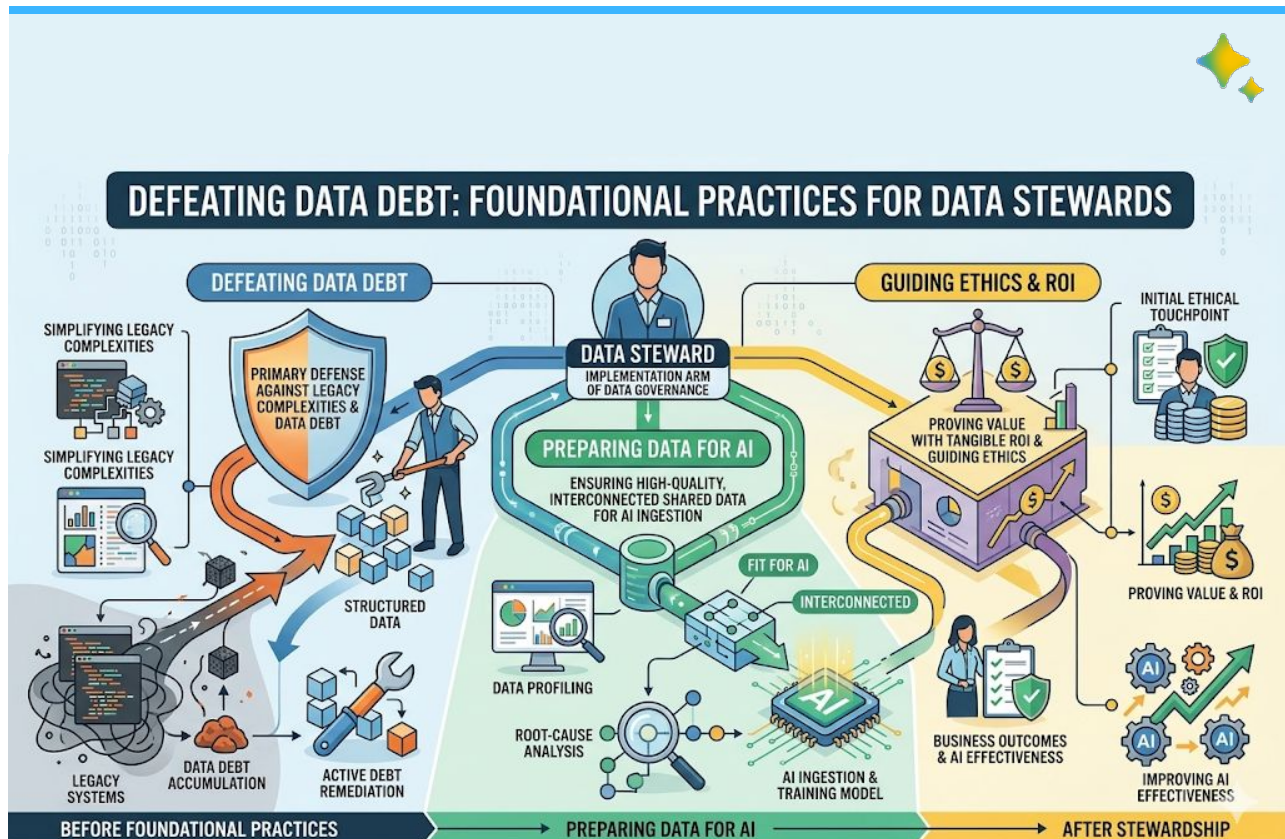
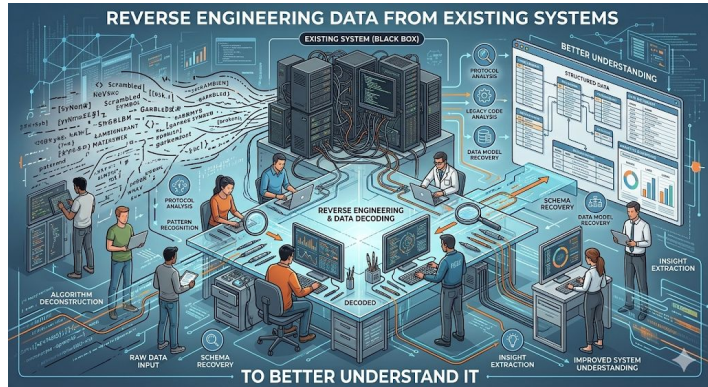


Compliment Architectural (Re)Developmental Opportunities

- Part of a larger effort
- Use AI as an "expert companion" to overcome organizational challenges and eliminate legacy complexities

Act as the

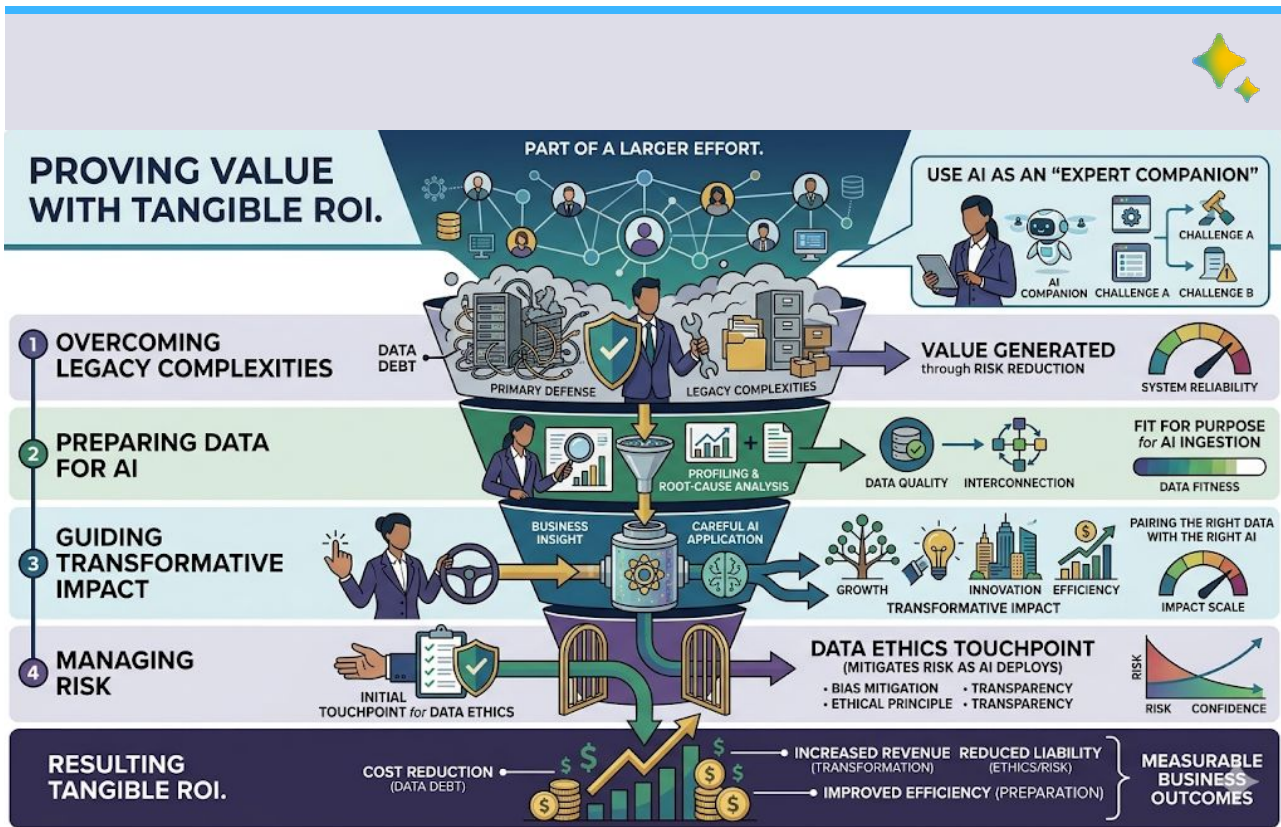
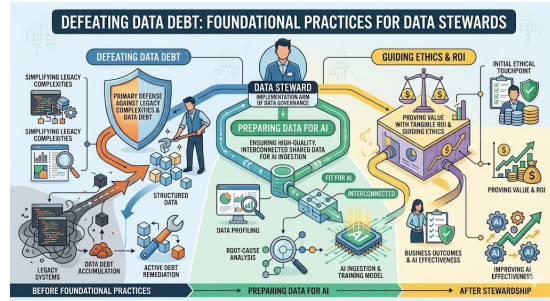
- Implementation arm of data governance,
- Fighting "data debt" and
- Ensuring that the organization's data is high-quality, interconnected, and fit for AI ingestion



Understanding the Cadence/Role of Foundational Data Practices Supporting Organizational Operations



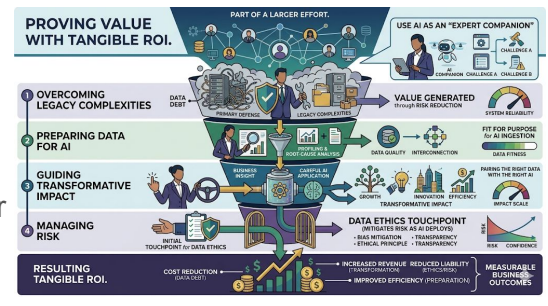
- Part of a larger effort
- Use AI as an "expert companion" to assist data stewards with specific organizational challenges
- **Defeating Data Debt:** Foundational practices allow data stewards, who serve as the "implementation arm of data governance," to act as the primary defense against legacy complexities and data debt.
- **Preparing Data for AI:** By applying these practices in their organizational operations, stewards can perform profiling and root-cause analysis to ensure that shared data is high-quality, interconnected, and fit for AI ingestion.
- **Guiding Ethics and ROI:** Understanding the cadence of these practices helps stewards serve as the initial touchpoint for data ethics, prove value with tangible ROI, and improve AI effectiveness from an organization-wide perspective.



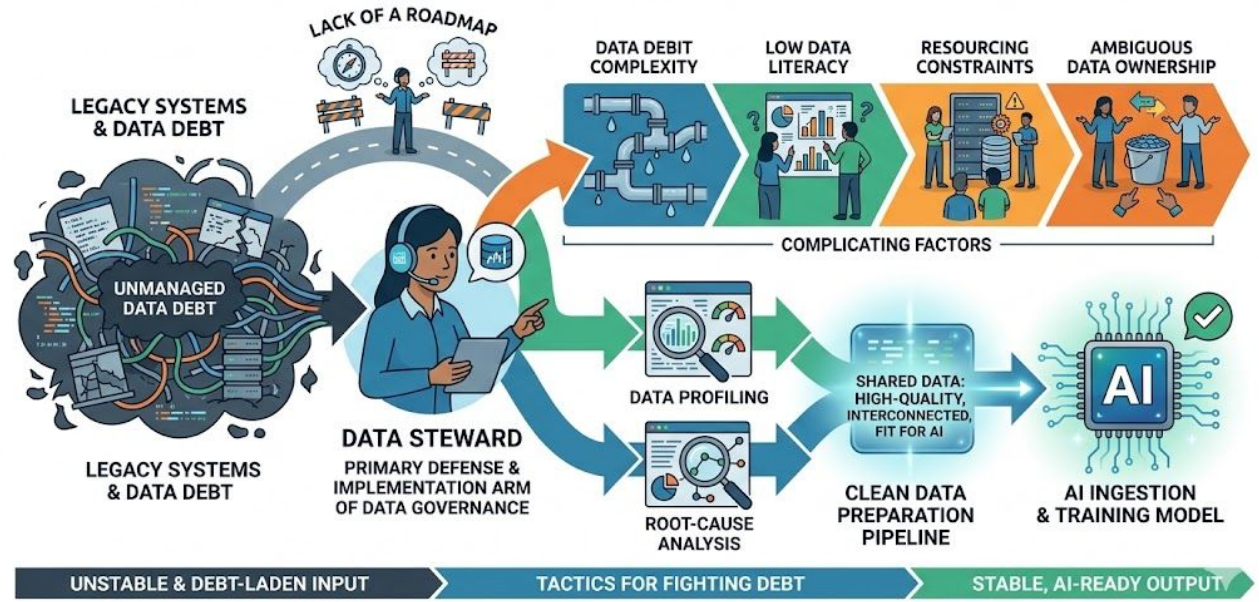
Proving Value With Tangible ROI



- Part of a larger effort
- Use AI as an "expert companion" to assist data stewards with specific organizational challenges
- **Guiding Transformative Impact:** Stewards possess the necessary business insight to pair the right data with the careful application of AI, which guides organizations toward transformative impact.
- **Overcoming Legacy Complexities:** They provide value by serving as the primary defense against "data debt" and legacy complexities that cause organizational plans to fail.
- **Preparing Data for AI:** Stewards perform profiling and root-cause analysis to ensure data is interconnected, high-quality, and "fit for purpose" for AI ingestion.
- **Managing Risk:** They act as the initial touchpoint for data ethics, which mitigates risk as organizations deploy AI



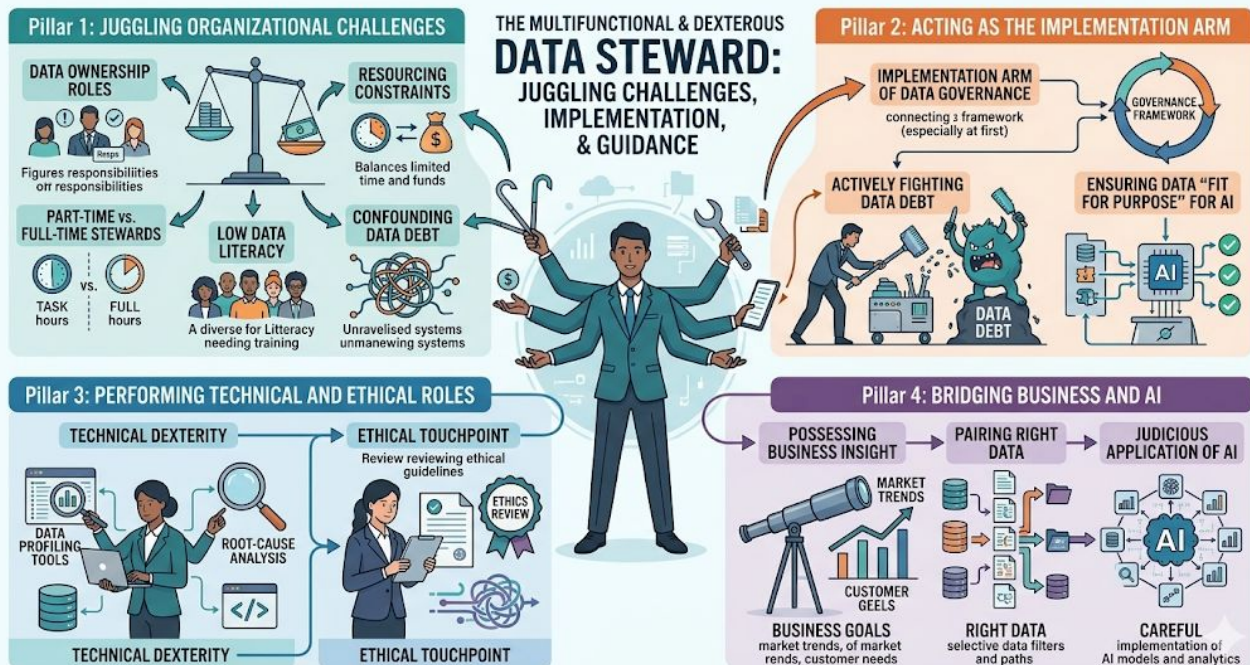
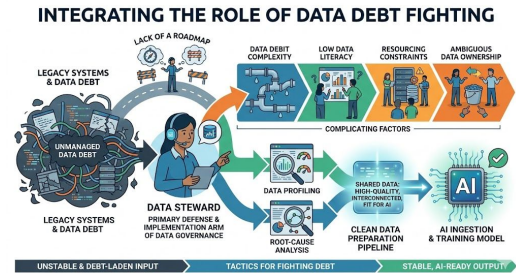
INTEGRATING THE ROLE OF DATA DEBT FIGHTING



Integrating the Role of Fighting Data Debt



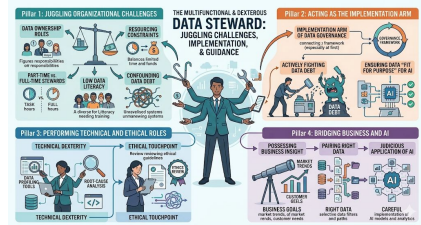
- Part of a larger effort
- Use AI as an "expert companion" to assist data stewards with specific organizational challenges
- **The Primary Defense:** As the implementation arm of data governance, data stewards serve as the primary defense against data debt.
- **Lack of a Roadmap:** Many stewards currently lack a roadmap to eliminate the data debt and legacy complexities that cause even the best organizational plans to fail.
- **Complicating Factors:** Organizational challenges are often worsened by confounding "data debt" complexities, which sit alongside other hurdles like low data literacy, resourcing constraints, and ambiguous data ownership.
- **Tactics for Fighting Debt:** Stewards actively combat this data debt by performing profiling and root-cause analysis. Taking these steps ensures that shared data is high-quality, interconnected, and actually fit for AI ingestion.



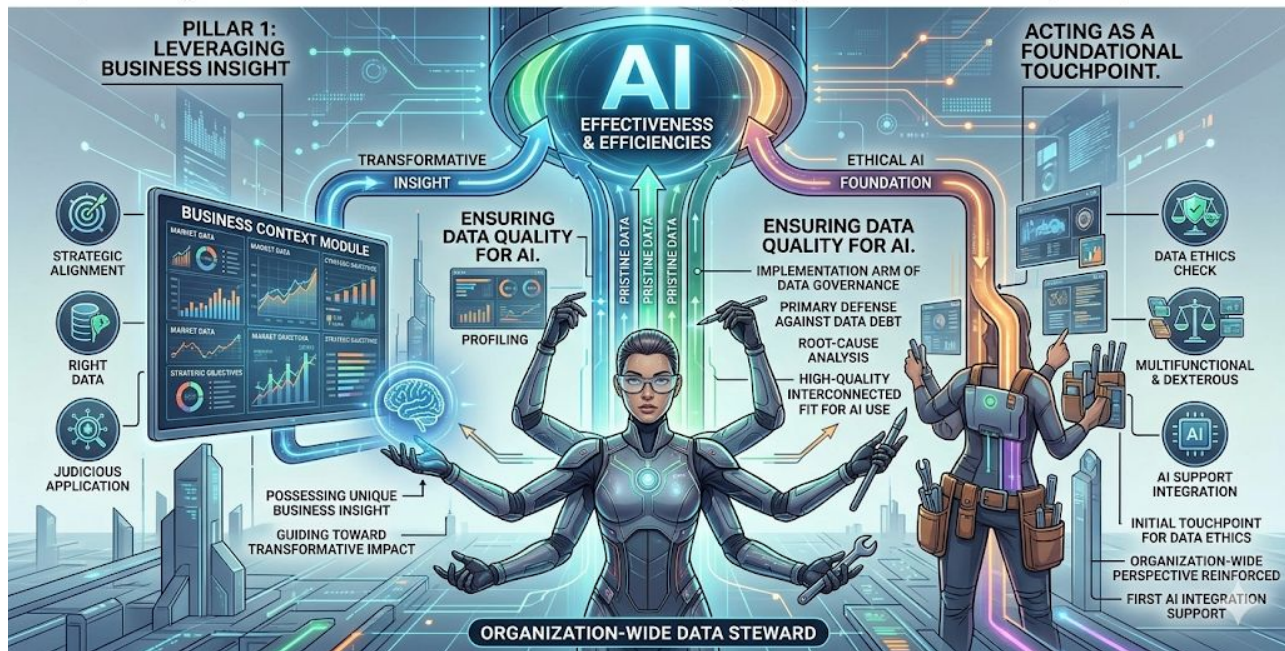
Comprehending How Stewards Need To Be Multifunctional and Dexterous (Especially at First)



- Part of a larger effort
- Use AI as an "expert companion" to assist data stewards with specific organizational challenges
- **Juggling Multiple Organizational Challenges:** Stewards must navigate complicated issues, including the specific role of data ownership, resourcing constraints (full-time vs. part-time), generally low data literacy, and confounding data debt complexities.
- **Acting as the Implementation Arm:** They serve as the implementation arm of data governance, which requires them to actively fight data debt and ensure data is "fit for purpose" for AI ingestion.
- **Performing Technical and Ethical Roles:** Their duties require them to be dexterous enough to perform technical tasks like profiling and root-cause analysis, while simultaneously serving as the initial organizational touchpoint for data ethics.
- **Bridging Business and AI:** They are uniquely positioned to guide transformative impact because they must possess the business insight necessary to pair the right data with the judicious application of AI.



Improving AI effectiveness/efficiencies using organization-wide perspectives

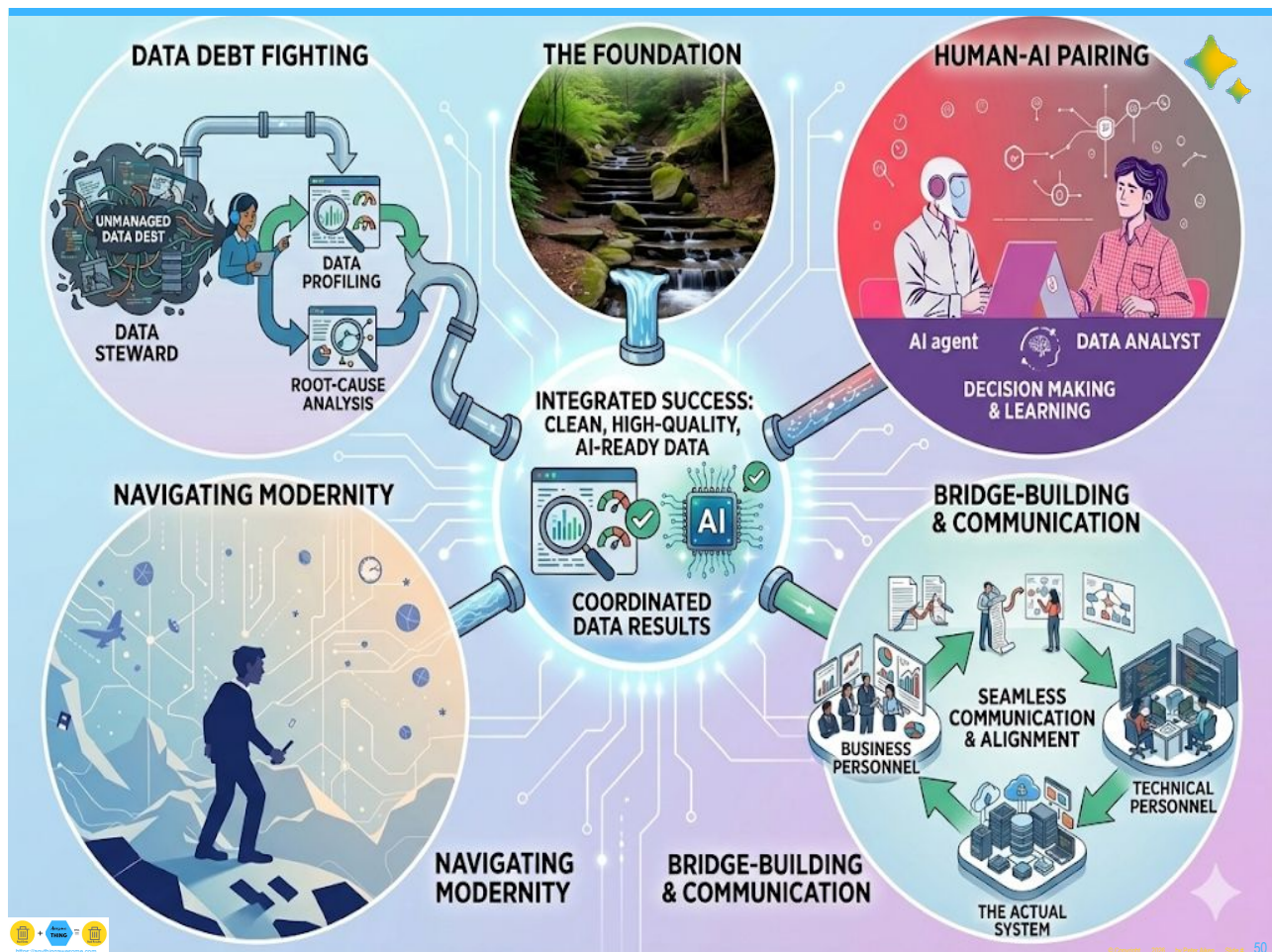
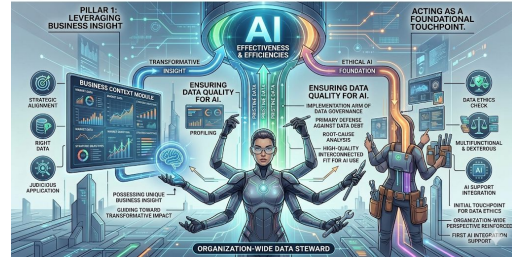


Improving AI Effectiveness/Efficiencies Using Organization-Wide Perspectives



- Part of a larger effort
- Use AI as an "expert companion" to assist data stewards with specific organizational challenges
- **Leveraging Business Insight:** Stewards are uniquely positioned to guide organizations toward transformative impact because they possess the business insight necessary to pair the right data with the judicious application of AI.
- **Ensuring Data Quality for AI:** As the implementation arm of data governance, stewards serve as the primary defense against data debt. By conducting profiling and root-cause analysis, they ensure that shared data across the organization is high-quality, interconnected, and fit for AI use.
- **Acting as a Foundational Touchpoint:** Their organization-wide perspective is reinforced by their role as the initial touchpoint for data ethics and their need to be multifunctional and dexterous, especially when first integrating AI support

Improving AI effectiveness/efficiencies using organization-wide perspectives



Program overview



- AI Reality Check?
 - Barriers to entry
 - Not “fit for purpose”
 - Data debt
- Specifics (part of a larger effort)
 - Compliment architectural (re)developmental opportunities
 - Understanding the cadence/role of foundational data practices supporting organizational operations
 - Proving value with tangible ROI
 - Improving AI effectiveness/efficiencies using organization-wide perspectives
 - Comprehending how stewards need to be multifunctional and dexterous, especially at first
 - Integrating the role of data debt fighting
- Practical Uses
 - Vision/approach cohesion
 - Build Multifunctional Dexterity
 - Ethical support
- Take aways → Q&A

Data Stewards: Conquering Data Debt in the Age of AI



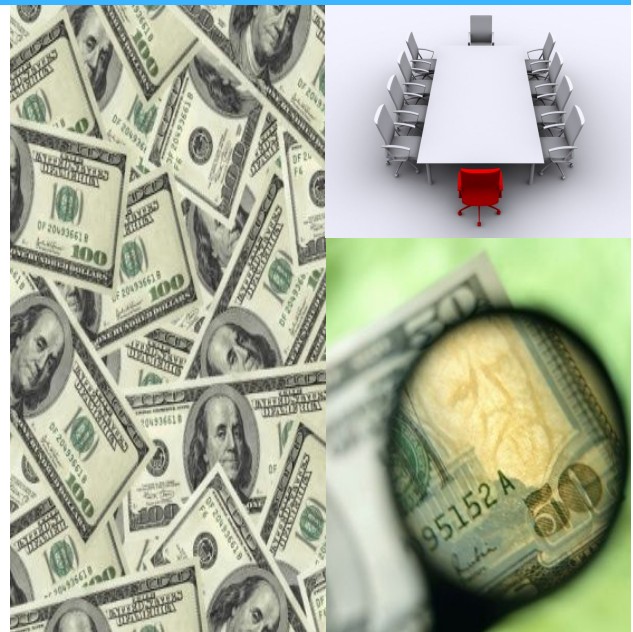
<https://creativecommons.org/licenses/by-nc-sa/4.0/>



© Copyright 2020 by Peter Allen Slide 51

Corporate Governance

- "Corporate governance - which can be defined narrowly as the relationship of a company to its shareholders or, more broadly, as its relationship to society....",
Financial Times, 1997.
- "Corporate governance is about promoting corporate fairness, transparency and accountability"
James Wolfensohn, World Bank,
President Financial Times, June 1999.
- “Corporate governance deals with the ways in which suppliers of finance to corporations assure themselves of getting a return on their investment”,
The Journal of Finance, Shleifer and Vishny, 1997.



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

IT Governance

- "Putting structure around how organizations **align IT strategy with business strategy**, ensuring that companies stay on track to achieve their strategies and goals, and implementing good ways to measure IT's performance.
- It makes sure that all stakeholders' interests are taken into account and that processes provide **measurable results**.
- Framework should answer some **key questions**, such as how the IT department is functioning overall, what key metrics management needs and what return IT is giving back to the business from the investment it's making." *CIO Magazine (May 2007)*

IT Governance Institute, 5 areas of focus:

- **Strategic Alignment**
- **Value Delivery**
- **Resource Management**
- **Risk Management**
- **Performance Measures**



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**



What is Data Governance?

Managing Data with Guidance

Go Ask
Anyone!™

*Would
you
want
your
sole,
non-
depletable,
non-
degrading,
durable,
strategic
asset
managed
without
guidance?*



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

55

What is Data Governance?

Managing Data Decisions with Guidance

Go Ask
Anyone!™

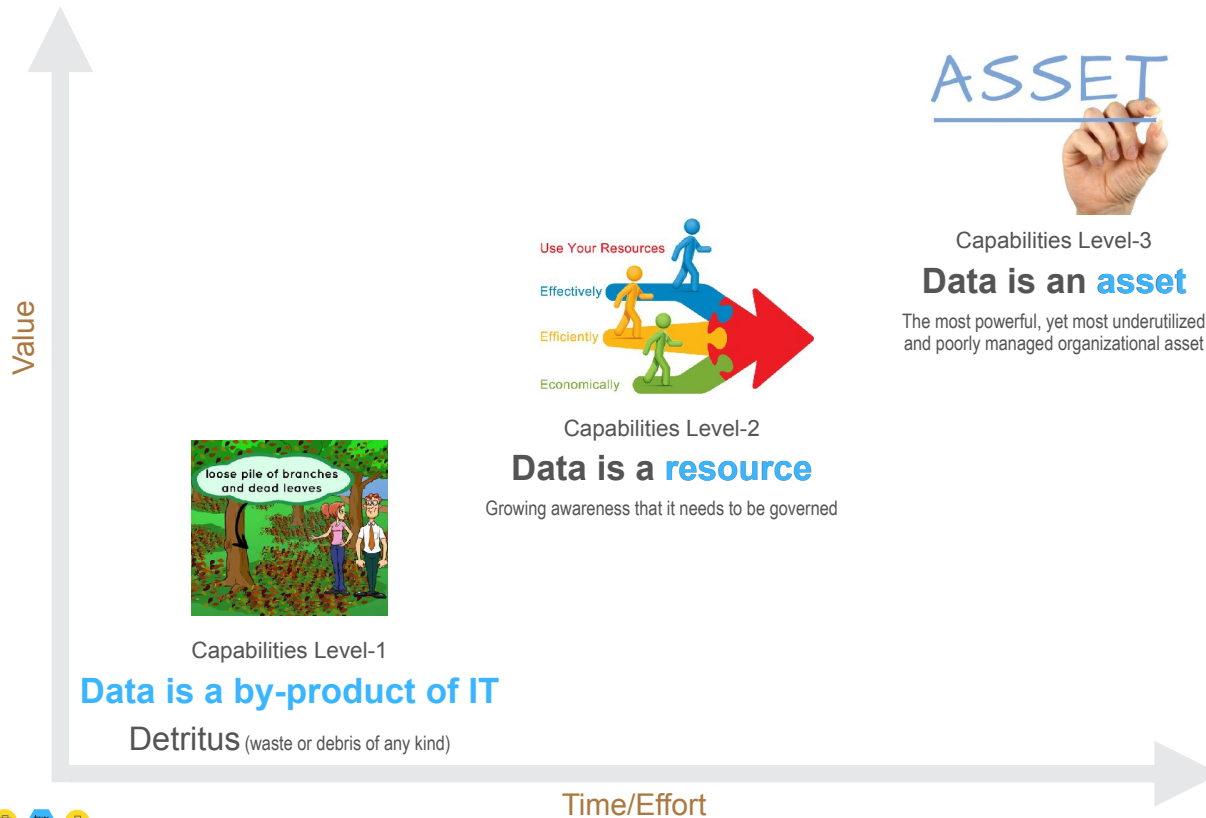
*Would
you
want
your
sole,
non-
depletable,
non-
degrading,
durable,
strategic
asset
managed
without
guidance?*



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

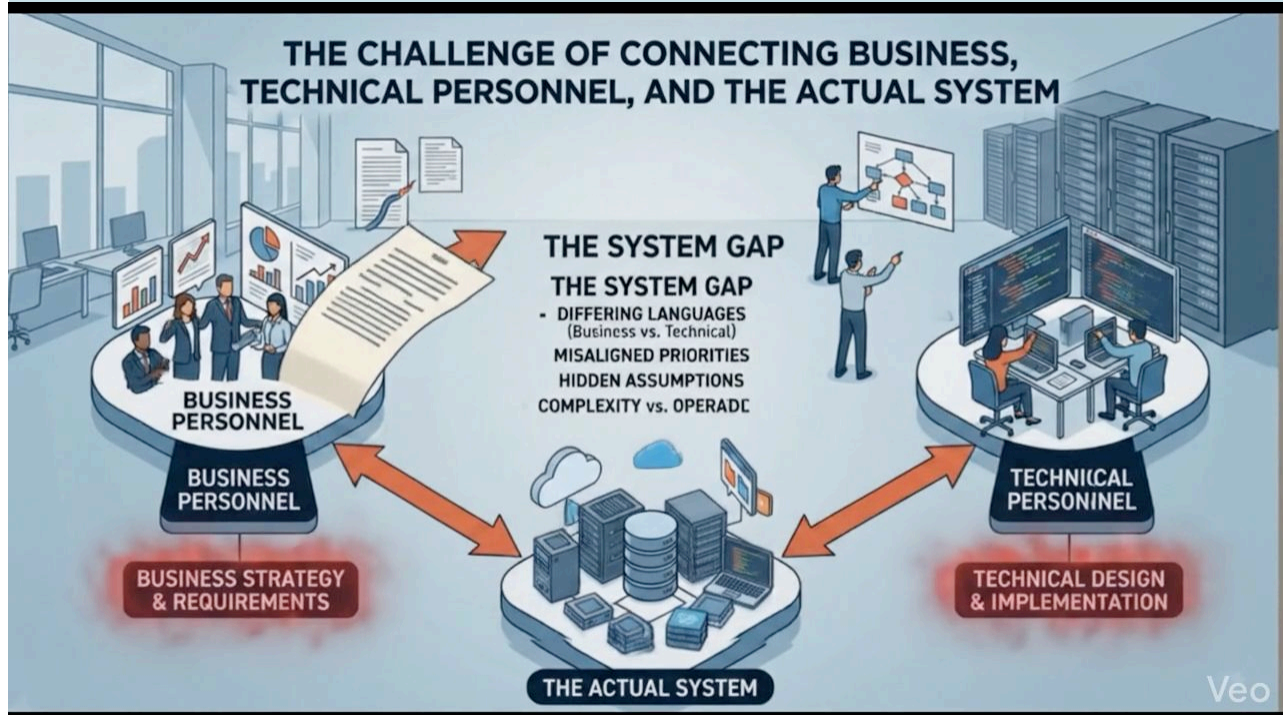
56

Connecting Data Decisions to Value

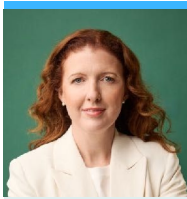


<https://anyringwallstone.com>

Achieve Shared Understanding (Minimally Among)



<https://anyringwallstone.com>

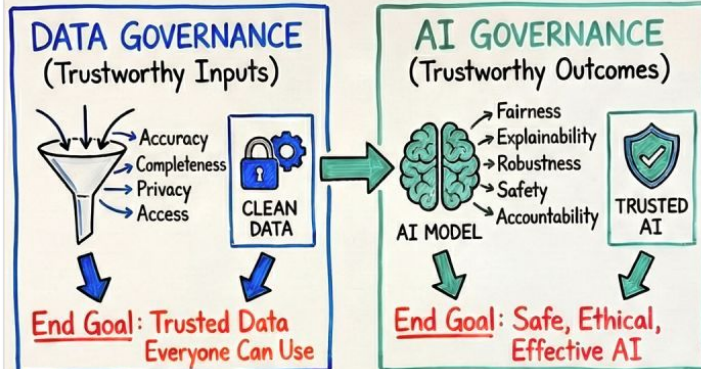


Data Governance Contrasted with AI Governance

https://www.linkedin.com/posts/clarekitching_the-fastest-way-to-understand-ai-governance-activity-7417117020757970944-k8Be?utm_source=share&utm_medium=member_desktop&rcm=ACoAAABQvAYBeDJR2H5Za01_13GOvBvARazH6fs

DATA GOVERNANCE vs AI GOVERNANCE

INPUTS → OUTCOMES



- **Stakeholders:** Stewards, Architects, IT, Security
- **Risks:** Breaches, Poor Quality, Bad Decisions
- **Maturity:** Established, Operational Hygiene
- **Pitfalls:** Admin Focus, Unclear Ownership

- **Stakeholders:** Legal, Risk, Ethics, Product Teams
- **Risks:** Harmful Bias, Drift, Opaque Models, Reputational Damage
- **Maturity:** Emerging, Board/Regulator Driven
- **Pitfalls:** One-off Audits, Ignoring Impact

Keep the proper focus

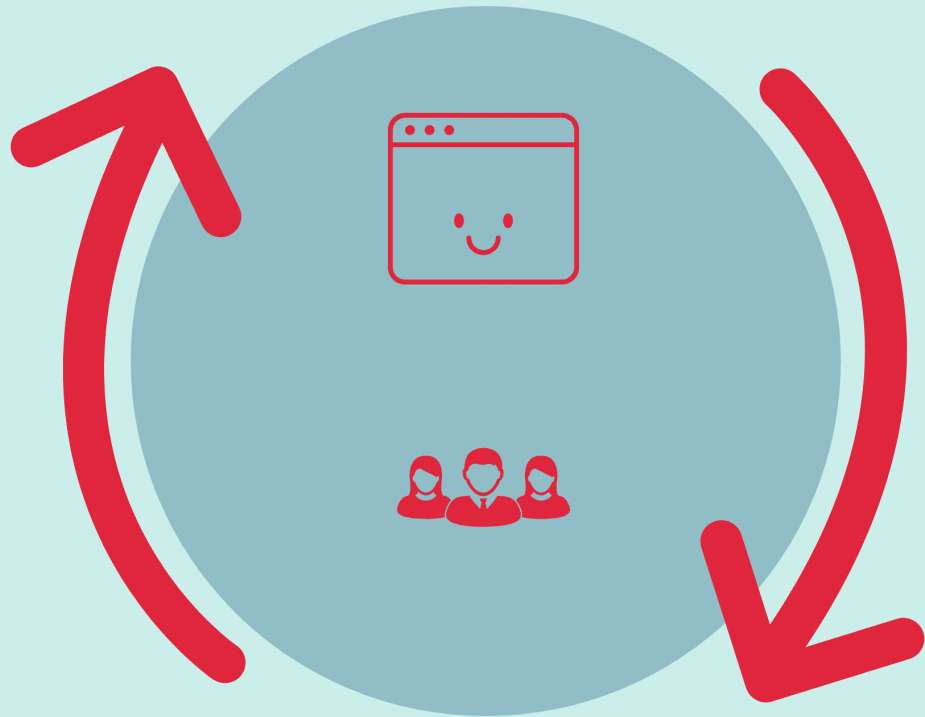
- Wrong question:
 - How should we govern all this data?
- Right question:
 - Should we include this data item within the scope of our current data governance practices?



- Regardless of the decision, document why!

(Especially at first) stewards require a united purpose

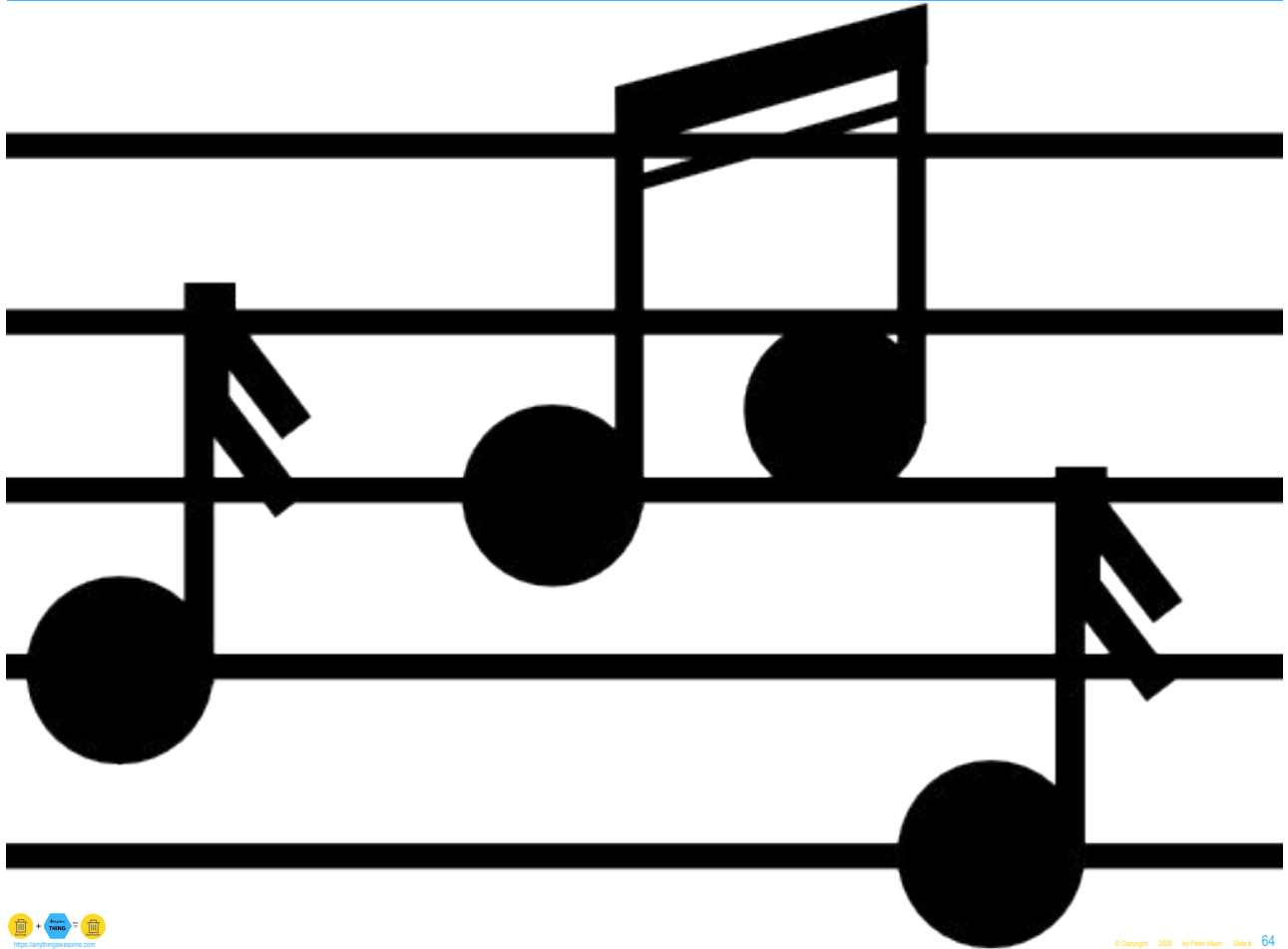
- Enable the organization to better use its data in support of the mission



Transform This



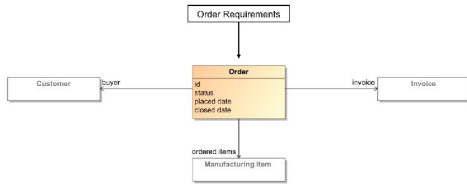
To This



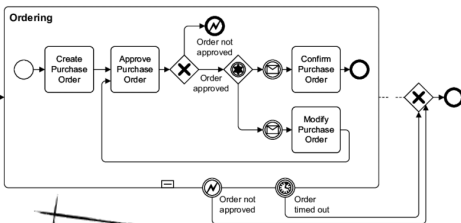


What does it need to work? A CRUD Matrix

Context Model

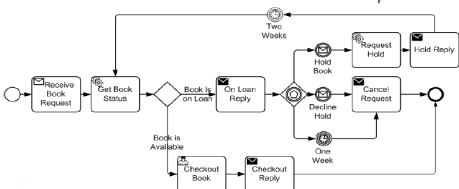


Process A

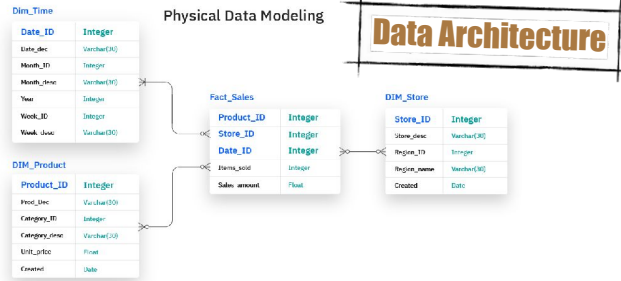


Process Architecture

Process B



Data Model



	Data Item 1	Data Item 2	Data Item 3	Data Item 4	Controlled Vocabulary
Process A	Creates	Reads	Updates		
Process B	Reads	Creates	Deletes		
Process C	Updates	Updates	Creates	Updates	
Process D	Deletes		Reads	Creates	
Controlled Vocabulary					

CRUD Matrix

CRUD is an acronym that refers to the following actions on an object:

- Create - to create and store new data
- Read - to retrieve and read data
- Update - to change or modify then store the data.
- Delete - to delete or remove the data



Data 'Owners' must be Process Owners

- Data always belongs to the organization
- Shorthand for Data Requirements Owners
- Owning the requirements includes:
 - Gateway Requirements for data to successfully enter your documented enterprise process
 - Processing Requirements for data to increase itself, or something else tangible, in value
 - Downstream Requirements for data produced by your operation to be successfully ingested in downstream operations
- Specifications for this portion of the organization to function correctly—**process owners**



DIMENSIONS OF DATA OWNERSHIP: THE DATA OWNER AS A STRATEGIC FIDUCIARY



ADVICE FOR MENTEES: VALUE-FOCUSED OWNERSHIP



Dimensions of Data Ownership



Executive Summary: A Data Owner is a Strategic Fiduciary. They manage the data on behalf of the organization to ensure it remains an asset and does not become a liability.

The Data Owner provides the 'Why' and the 'What'.

1. Decision Rights & Approval Authority

A business leader (often a Subject Matter Expert) who has the ultimate **approval authority** for decisions about data within their domain. This includes:

- Access Rights: Deciding who can see, use, and update the data.
- Policy Enforcement: Ensuring the organization actually follows the rules they've documented.
- Quality Standards: Setting the "Fitness for Purpose" levels. If the data isn't good enough for the process, the Owner is the one who decides it's a "fail".

2. Business Terminology & Semantic Consistency

Ownership isn't just about processes; it's about **meaning**. The Data Owner owns the "Business Glossary" for their domain. If Finance and Sales have two different definitions for "Customer Revenue," the Data Owner is the one who must broker that agreement to ensure a "[Single Version of the Truth](#)".

3. Risk Management & Compliance

In today's regulatory environment, "Owning Requirements" must include owning the **Risk Profile**.

- **Regulatory Categorization:** Identifying PII or financially sensitive data.
- **Data Retention:** Deciding when data must be archived/destroyed to meet legal and business mandates.
- **Auditability:** Being able to prove to a regulator that requirements were actually met.

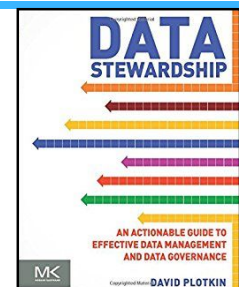
4. Valuation and Monetization

As an executive-level concept, owners should understand the **economic value** of their data. This isn't just about processing; it's about treating data as an asset on the balance sheet. An Owner should be able to articulate how their data subset increases revenue or reduces operational costs.



Data Steward

- **Business data steward**
 - Manage from the perspective of business elements (i.e. business definitions and data quality)
- **Technical data steward**
 - Focus on the use of data by systems and models (i.e. code operation)
- **Project data steward**
 - Gather definitions, quality rules and issues for referral to business/technical stewards
- **Domain data steward**
 - Manage data/metadata required across multiple business areas (i.e. customer data)
- **Operational data steward**
 - Directly input data or instruct those who do; aid business stewards identifying root cause and addressing issues
- **Metadata Data Steward**
 - Manage metadata as an asset
- **Legacy Data Steward**
 - Manage legacy data as an asset
- **Data steward auditor**
 - Ensures compliance with data guidance
- **Data steward manager**
 - Planning, organizing, leading and controlling



(list adapted from Plotkin, 2014)



Business Data Steward vs Technical Data Steward

by DQOps - data quality platform: <https://dqops.com/>

A data steward is a guardian of an organization's data, ensuring its accuracy, quality, and appropriate use. The distinction between business and technical data stewards promotes a holistic approach to data governance, integrating strategic goals with technical implementation.



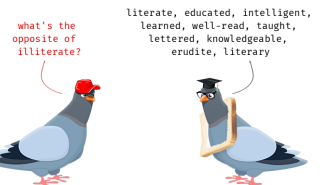
If you found this helpful, visit our website <https://dqops.com/> and download our free eBook on data quality



2020 by Peter Allen Slide 6 71



- Always pick one full time over 10% of 10 individuals
- Generally low data literacy rates among knowledge workers and management



2020 by Peter Allen Slide 6 72

Give me an argument for full time instead of part time data stewards



While part-time data stewards might seem cost-effective initially, employing full-time data stewards offers significant long-term advantages that ultimately contribute to a stronger data governance program and better business outcomes. Here's a breakdown of the argument for full-time data stewards:

1. Dedicated Focus and Ownership:

- Full-time stewards can fully immerse themselves in data governance activities. Part-time stewards, juggling other responsibilities, often struggle to prioritize data governance, leading to inconsistencies and delays.
- Dedicated time allows for proactive data management. Full-time stewards can anticipate issues, implement preventative measures, and continuously improve data quality, rather than just reacting to problems.
- Clear ownership fosters accountability. With data governance as their primary responsibility, full-time stewards are more likely to take ownership and be held accountable for data quality and compliance.

2. Improved Data Quality and Consistency:

- Consistent attention leads to better data quality. Full-time stewards have the time to thoroughly investigate data issues, identify root causes, and implement effective solutions.
- Full-time stewards can establish and enforce data standards and policies more effectively. This ensures consistency across the organization and reduces data silos.
- Proactive data monitoring and cleansing are more feasible. Full-time stewards can regularly monitor data quality and perform necessary cleansing activities, preventing data degradation over time.

3. Enhanced Collaboration and Communication:

- Full-time stewards are more readily available for collaboration. They can build stronger relationships with data users across the organization, fostering a culture of data awareness and responsibility.

- Dedicated time allows for effective communication and training. Full-time stewards can develop and deliver training programs on data governance policies and best practices, ensuring consistent understanding and compliance.
- Improved communication facilitates faster issue resolution. Full-time stewards can quickly address data-related issues and provide timely support to data users.

4. Increased Efficiency and ROI:

- Investing in full-time stewards can lead to greater efficiency in data-related processes. Improved data quality and accessibility can streamline operations and reduce costs associated with data errors and rework.
- Better data governance contributes to better decision-making. Accurate and reliable data empowers informed decision-making, leading to improved business outcomes and increased ROI.
- Reduced risk of data-related compliance issues. Full-time stewards can ensure compliance with relevant regulations, minimizing the risk of fines and reputational damage.

5. Long-Term Strategic Value:

- Full-time stewards can contribute to the development of a long-term data strategy. They have the time to understand the organization's business needs and align data governance efforts with strategic objectives.
- Dedicated focus allows for continuous improvement of the data governance program. Full-time stewards can identify areas for improvement and implement changes to enhance the effectiveness of the program.
- Investing in full-time stewards demonstrates a commitment to data as a strategic asset. This can attract and retain talent, as well as enhance the organization's reputation.

In conclusion, while the initial cost of hiring full-time data stewards might be higher, the long-term benefits in terms of improved data quality, enhanced efficiency, reduced risk, and better decision-making far outweigh the costs. Investing in full-time data stewards is a strategic investment in the organization's future.



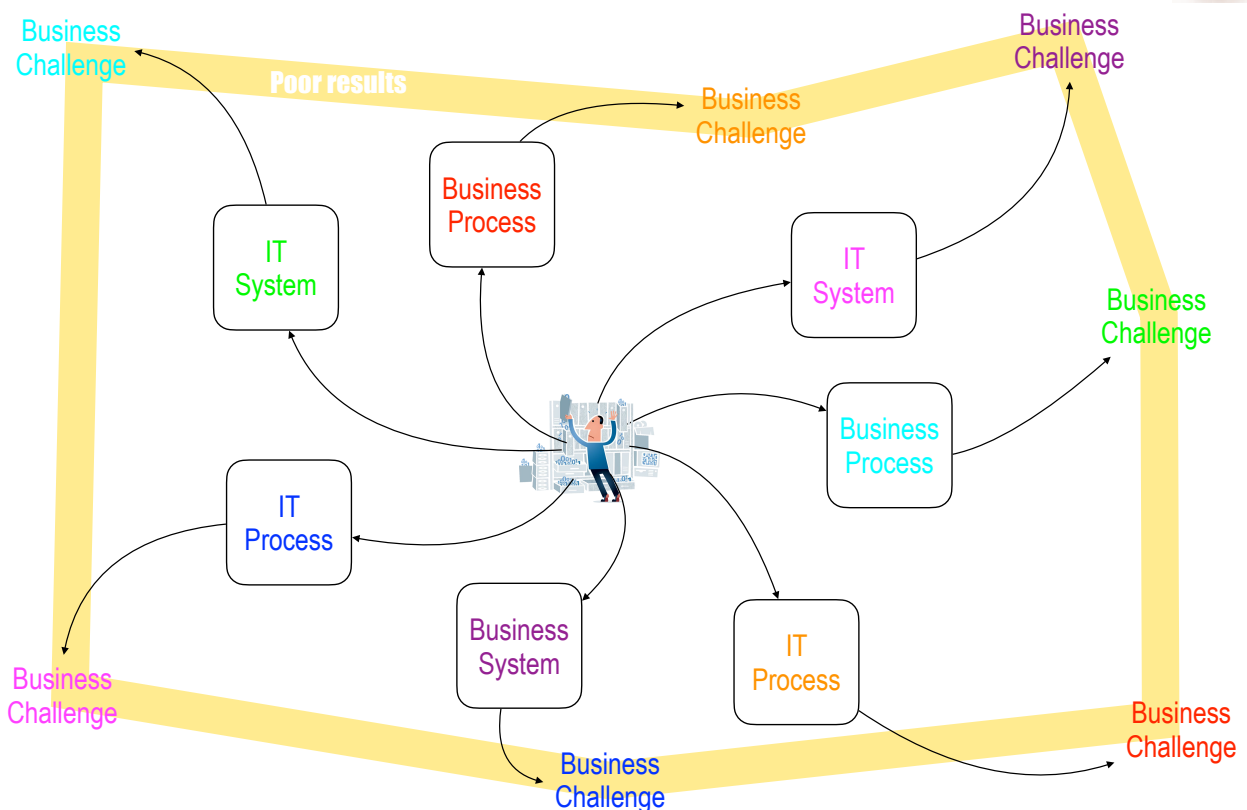
Reducing data debt in a manner that enables better AI ROI, requires multifunctional and dexterous data stewards



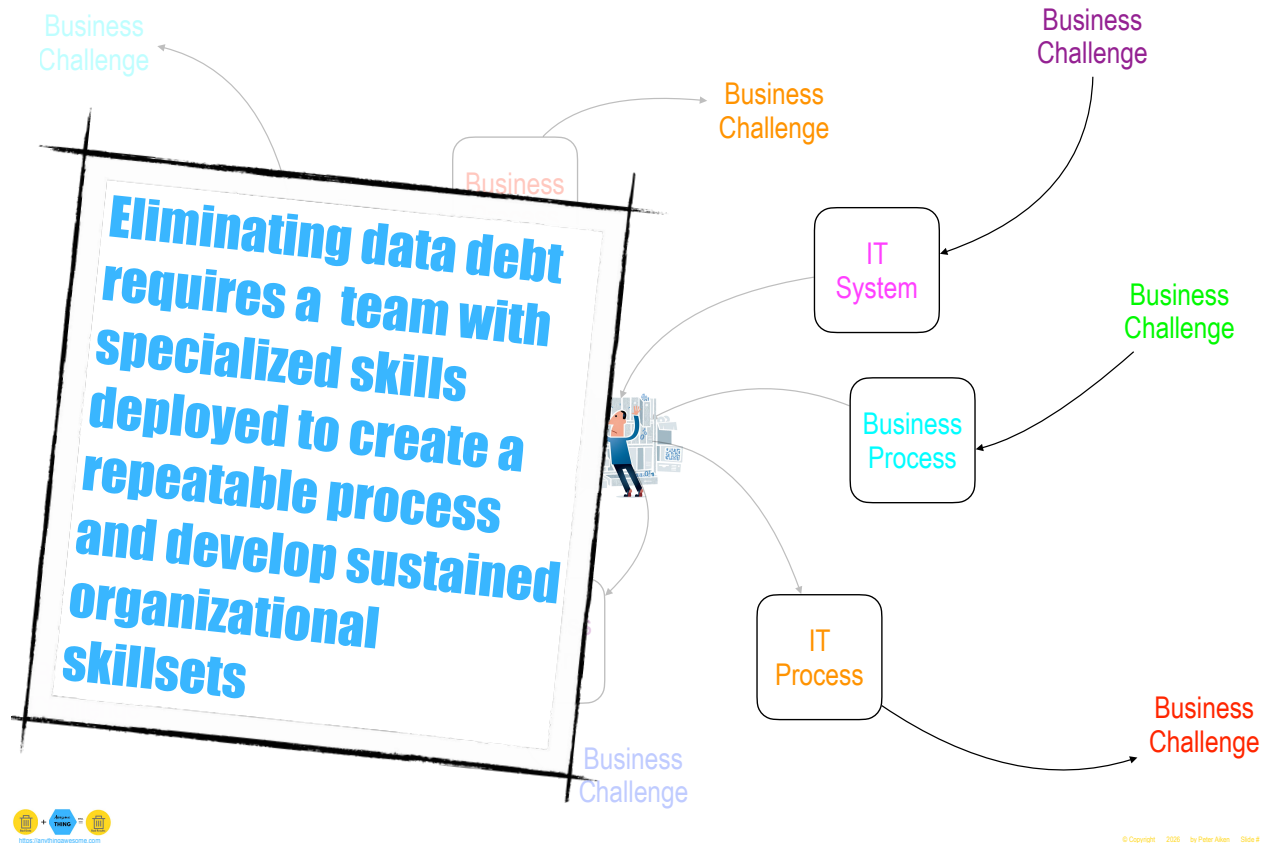
Poor data manifests as multifaceted organizational challenges



Root cause analysis is part of data stewardship



Consistency Encourages Quality Analysis



© Copyright 2020 by Peter Allen Slide # 77

From Cost Centers to Value Engines: Proactive Stewardship

- It is not all about data debt
 - Modern Data Stewards also serve as Strategic Fiduciaries who directly enable revenue generation.
 - #1 Focus: Eliminate "data hunting", allowing data scientists and product teams to launch ahead of schedule
- Enabling Hyper-Personalization
 - Broker definitions between silos and manage Master Data to ensure largest customer profiles are accurate-improving conversion rates
- Monetization & Valuation
 - Data Owners must be able to articulate precisely how their data subset increases revenue
 - Real-World Example: Just as airlines unlock billions in market value by monetizing loyalty data, stewards ensure data assets are high-quality enough to be leveraged for new revenue streams



© Copyright 2020 by Peter Allen Slide # 78

Write a Data Governance Charter for a Small Non-Profit That Rescues Horses



1. Introduction

[Non-Profit Name] recognizes that data is a critical asset for achieving our mission of rescuing and rehabilitating horses. Effective data governance is essential to ensure data quality, integrity, security, and accessibility, enabling informed decision-making and efficient operations. This charter establishes the framework for data governance at [Non-Profit Name].

2. Mission and Objectives

The mission of this Data Governance program is to:

- Ensure data is accurate, reliable, and consistent for effective program management and reporting.
- Protect sensitive data related to donors, volunteers, staff, and the horses in our care.
- Optimize data utilization for fundraising, outreach, and operational efficiency.
- Promote data literacy and a data-driven culture within the organization.
- Comply with all applicable legal and regulatory requirements related to data privacy and security.

3. Scope

This charter applies to all data created, collected, processed, and stored by [Non-Profit Name], including but not limited to:

- Donor information
- Volunteer information
- Horse records (medical, behavioral, training)
- Financial data
- Website analytics
- Social media data
- Program data (adoptions, rescues, etc.)

4. Guiding Principles

The Data Governance program will be guided by the following principles:

- **Accountability:** Clearly defined roles and responsibilities for data management.
- **Transparency:** Open communication and access to data governance policies and procedures.
- **Integrity:** Maintaining the accuracy, completeness, and consistency of data.
- **Security:** Protecting data from unauthorized access, use, or disclosure.
- **Accessibility:** Ensuring data is readily available to authorized users.
- **Compliance:** Adhering to all relevant legal and regulatory requirements.

5. Data Governance Roles and Responsibilities

Data Governance Council: Composed of key stakeholders (e.g., Executive Director, Program Manager, Development Director, Volunteer Coordinator). Responsible for:

- Setting data governance policies and standards.
- Approving major data-related initiatives.
- Monitoring data quality and compliance.
- Resolving data-related disputes.

Data Steward(s): Designated individuals responsible for managing specific data domains

- Donor Data Steward, Horse Records Steward). Responsibilities include: ensuring data quality within their domain.

- Implementing data governance policies.
- Training staff on data management procedures.
- **Data Custodian(s):** Individuals responsible for the technical management of data and systems (e.g., IT staff, database administrators). Responsibilities include:
 - Implementing data security measures.
 - Maintaining data infrastructure.
 - Ensuring data backup and recovery.
- **Data Users:** All staff and volunteers who interact with data. Responsible for:
 - Adhering to data governance policies and procedures.
 - Reporting data quality issues.
 - Using data responsibly and ethically.

6. Data Governance Processes

Data Quality Management: Implementing processes for data validation, cleansing, and enrichment.

- **Data Security and Privacy:** Establishing security measures to protect data from unauthorized access and use.
- **Data Access and Sharing:** Defining clear guidelines for data access and sharing, both internally and externally.
- **Data Retention and Disposal:** Developing policies for data retention and secure disposal of data.
- **Data Training and Awareness:** Providing training to staff and volunteers on data governance policies and best practices.

7. Data Governance Policies and Standards

The Data Governance program will develop and maintain policies and standards related to:

- Data quality
- Data security
- Data privacy
- Data access
- Data retention
- Data dictionary and metadata management

8. Review and Updates

This Data Governance Charter will be reviewed and updated annually or as needed to reflect changes in the organization's needs or regulatory environment.

9. Contact Information

For questions or concerns regarding this Data Governance Charter, please contact

[Designated Contact Person/Department].

This charter serves as a foundational document for data governance at [Non-Profit Name]. Its successful implementation will require ongoing commitment and collaboration from all members of the organization.



RACI (Responsible, Accountable, Consulted, and Informed)



Tasks	Maintenance supervisors	Maintenance analyst	Maintenance planner	Maintenance technician	Maintenance support	Rel specialist	CMMS project engineer
Inputting failure data	A	C	I	R		C	C
Work order completion	R	C	C	C	A	I	I
Work order closeout	C	R	C		I	I	A
QA of failure data input	C	R	I	C	I	C	A
Analyze failure reports	C	C	I	C	A	R	I
Maintenance strategy adjustments	C	I	I	C	A	R	R
Implementing new strategies	R	I	R	C	A	I	I

R = Responsible (also recommender)

Those who complete the task.[7] There is at least one role with a participation type of responsible, although others can be delegated to assist in the work required. (See also RASCI below for separately identifying those who participate in a supporting role.)

A = Accountable (also approver or final approving authority)

The one ultimately answerable for the correct and thorough completion of the deliverable or task, the one who ensures the prerequisites of the task are met and who delegates the work to those responsible.[7] In other words, an accountable must sign off (approve) work that responsible provides. There must be only one accountable specified for each task or deliverable.[8]

C = Consulted (sometimes consultant or counsel)

Those whose opinions are sought, typically subject-matter experts, and with whom there is two-way communication. [7]

I = Informed (also informee)

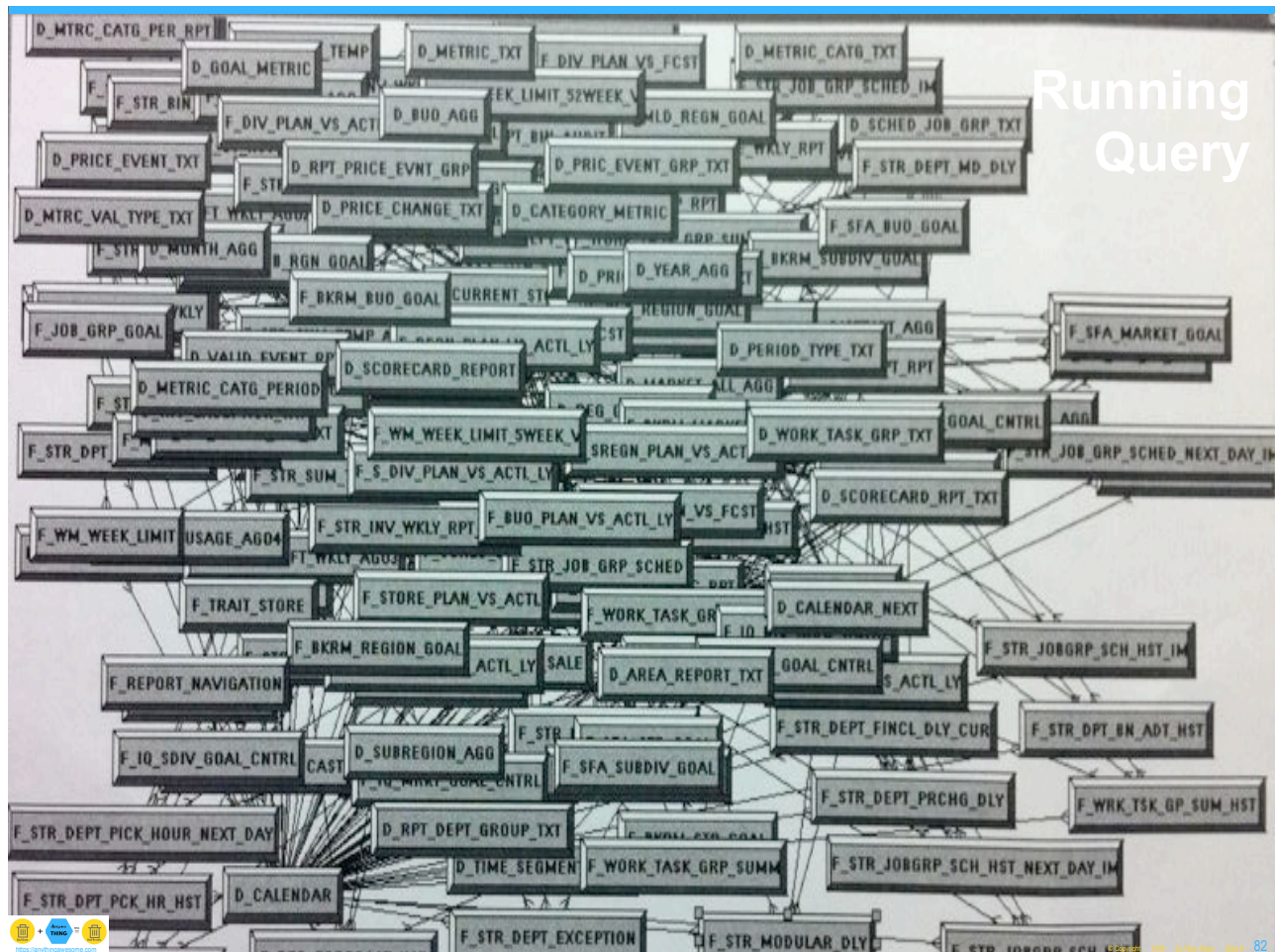
Those who are kept up-to-date on progress, often only on completion of the task or deliverable, and with whom there is just one-way communication.[7]

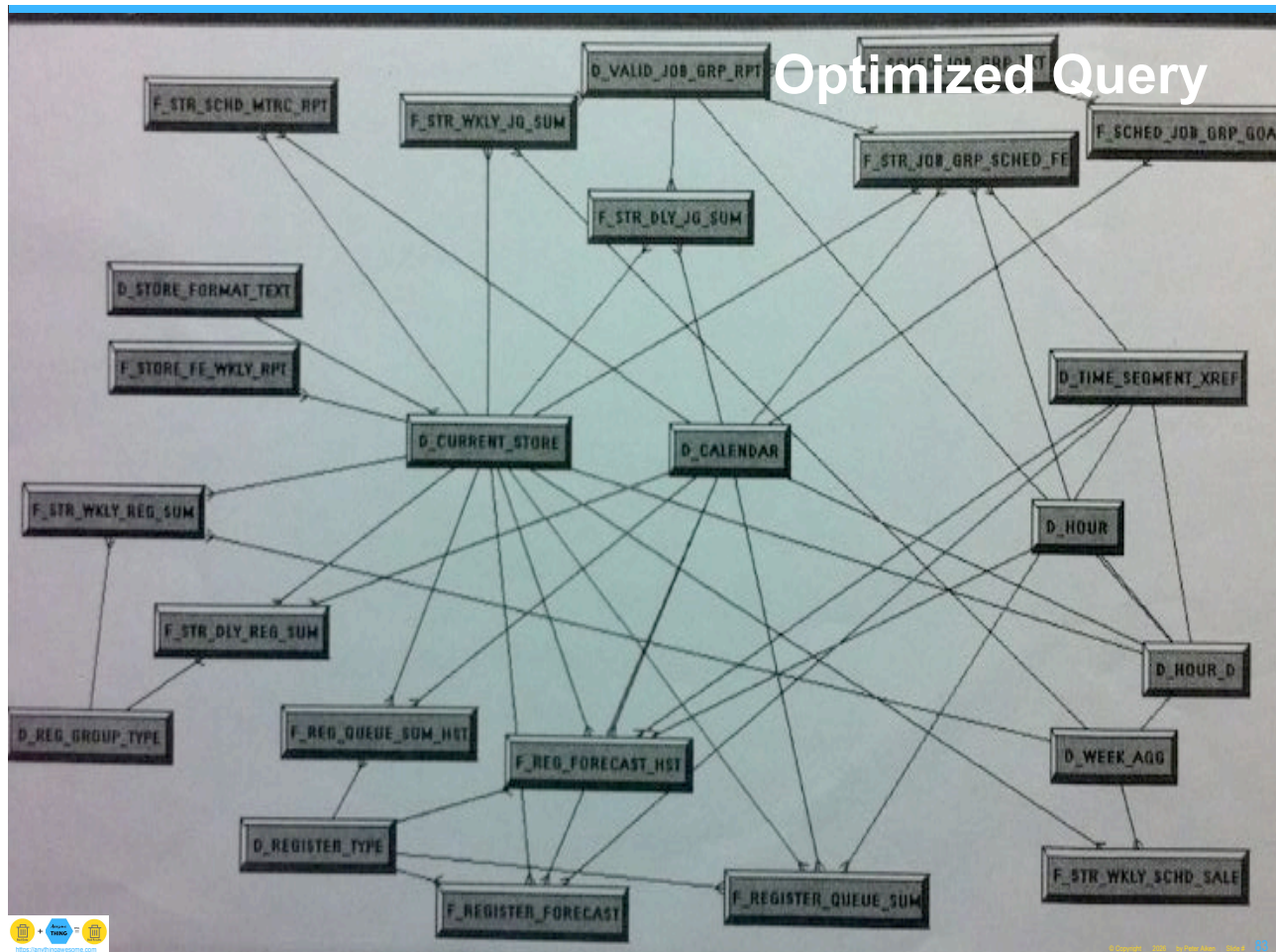


Create a RACI Matrix Illustrating Data Governance Roles Among the Business and IT Users and Data Stewards



Activity	Business Users	IT Users	Data Stewards
Data Strategy Development	C	I	A
Data Policy Creation & Approval	C	I	A
Data Architecture Design	C	A	C
Data Modeling	C	R	C
Data Definition & Business Glossary Management	C	I	A
Data Quality Rule Definition	C	C	R
Data Quality Monitoring & Reporting	I	R	A
Data Quality Issue Resolution	C	R	A
Metadata Management	I	R	A
Data Security Policy Implementation	I	A	C
Data Access Management	I	R	A
Data Privacy Compliance	I	A	C
Data Retention & Disposal Policy	C	I	A
Data Training & Awareness	I	C	R
Data Governance Tool Selection & Implementation	C	R	A
Data Issue Escalation & Resolution	C	R	A
Data Governance Program Evaluation	C	I	A





Data Footprints

- SQL Server
 - 47,000,000,000,000 bytes
 - Largest table 34 billion records 3.5 TBs
- Informix
 - 1,800,000,000 queries/day
 - 65,000,000 tables / 517,000 databases
- Teradata
 - 117 billion records
 - 23 TBs for one table
- DB2
 - 29,838,518,078 daily queries



Repeat 100s, thousands, millions of times ...



What do data stewards do in our organization?



- **Stewart**

- I'm a business subject matter expert, sitting in IT or LOB as a liaison between the two.
- I provide and distribute high quality data

- **Accountabilities**

Making data useful to the business

- Consistent use of data
- Promoting/achieving high data quality
- Resolving data integrity issues

- **Perspectives**

- Process and detail oriented with great organizational skills
- Prides himself on his creative resourcefulness, passion for quality and great interpersonal skills
- A 'de facto' steward because of deep industry expertise and understanding of his organization's data sources

- **Skills**

- 3+ years experience
- Understands data relationships, data process flows
- Proficient w/ Office, Power Pivot, SQL

- **Work Activities**

- Analyzes data for quality
- Reconciles data issues
- Identifies new data sources
- Actively analyzes data for 'semantic' quality
- Drives resolution of data issues
- Leads and/or participates in MDM / EIM / DQ initiatives
- Creates and maintains business metadata, references data values and meanings, and / or master data values and meanings

Tangibly (within scope)

1) Improve our data asset's value

2) Improve data's use



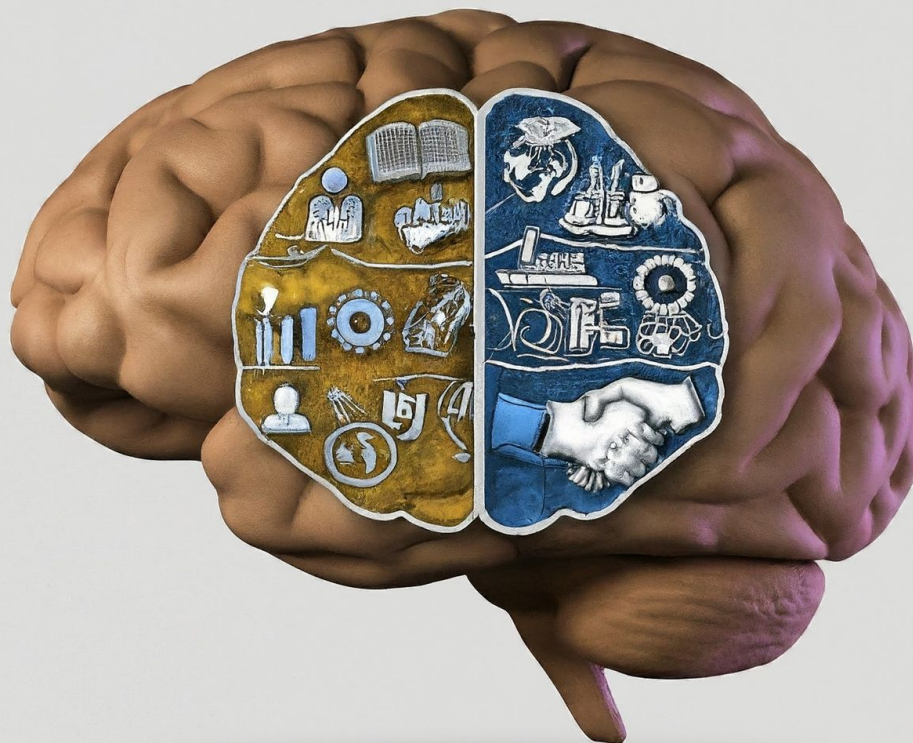


Compare Story Types



Technical	Business
Clean some data	Decrease the number of undeliverable targeted marketing ads
Reorganize the database	Increase the ability of the salesforce to perform their own analyses
Develop a taxonomy	Create a common vocabulary for the organization
Optimize a query	Shaved 1 second off a task that runs a billion times a day
Reverse engineer the legacy system	Understand: what was good about the old system so it can be formally preserved and, what was bad so it can be improved

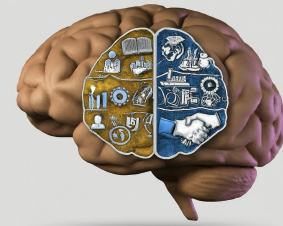
Knowledge, Skills and Abilities for a Data Steward



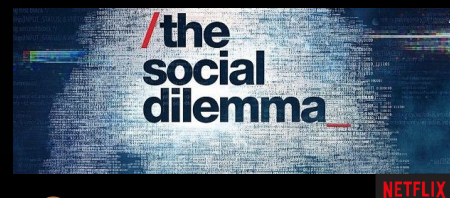
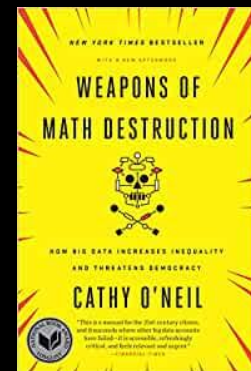
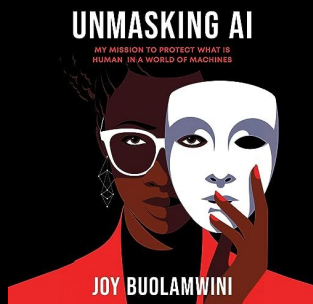
Knowledge, Skills and Abilities for a Data Steward

- In-depth knowledge of one or more data systems
- Knowledge of data standards
- Understanding of database design principles
- Understanding of private and sensitive data classification and protection
- Understanding of data interoperability challenges and solutions
- Ability to create and maintain metadata
- Ability to create and maintain business rules
- Ability to recognize and diagnose data quality issues
- Ability to develop data retention schedules
- Ability to identify and advocate for data quality and usability improvements
- Ability to effectively communicate technical information
- Ability to engage with data users to understand their needs

(Source: NCHRP Project 23-23 Implementing Data)



"Should I-thinking" must replace "Can I-thinking"



Data Ethics



Data Ethics Canvas

<p>Data sources</p> <p>Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.</p> <p>Is any personal data involved, or data that is otherwise sensitive?</p>	<p>Limitations in data sources</p> <p>Are there limitations that could influence your project's outcomes?</p> <ul style="list-style-type: none"> — bias in data collection, inclusion/exclusion, analysis, algorithms — gaps or omissions in data — provenance and data quality — other issues affecting decisions, such as team composition 	<p>Sharing data with others</p> <p>Are you going to be sharing data with other organisations? If so, who?</p> <p>Are you planning to publish any of the data? Under what conditions?</p>	<p>Ethical and legislative context</p> <p>What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?</p> <p>Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).</p>	<p>Rights around data sources</p> <p>Where did you get the data from? Is it produced by an organisation or collected directly from individuals?</p> <p>Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?</p>
<p>Your reason for using data</p> <p>What is your primary purpose for collecting and using data in this project?</p> <p>What are your main use cases? What is your business model?</p> <p>Are you making things better for society? How and for whom?</p> <p>Are you replacing another product or service as a result of this project?</p>	<p>Communicating your purpose</p> <p>Do people understand your purpose — especially people who the data is about or who are impacted by its use?</p> <p>How have you been communicating your purpose? Has this communication been clear?</p> <p>How are you ensuring more vulnerable individuals or groups understand?</p> <p>How are you ensuring more vulnerable individuals or groups understand?</p>	<p>Positive effects on people</p> <p>Which individuals, groups, demographics or organisations will be positively affected by this project? How?</p> <p>How are you measuring and communicating positive impact? How could you increase it?</p>	<p>Negative effects on people</p> <p>Who could be negatively affected by this project?</p> <p>Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?</p> <p>How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organisations using the data.</p>	<p>Minimising negative impact</p> <p>What steps can you take to minimise harm?</p> <p>How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure?</p> <p>How are you measuring, reporting and acting on potential negative impacts of your project?</p> <p>What benefits will these actions bring to your project?</p>
<p>Engaging with people</p> <p>How can people engage with you about the project?</p> <p>How can people correct information, appeal or request changes to the product/service? To what extent?</p> <p>Are appeal mechanisms reasonable and well understood?</p>	<p>Openness and transparency</p> <p>How open can you be about this project?</p> <p>Could you publish your methodology, metadata, datasets, code or impact measurements?</p> <p>Can you ask peers for feedback on the project?</p> <p>How will you communicate it internally?</p> <p>Will you publish your actions and answers to this canvas openly?</p>	<p>Ongoing implementation</p> <p>Are you routinely building in thoughts, ideas and considerations of people affected in your project? How?</p> <p>What information or training might be needed to help people understand data issues?</p> <p>Are systems, processes and resources available for responding to data issues that arise in the long-term?</p>	<p>Reviews and iterations</p> <p>How will ongoing data ethics issues be measured, monitored, discussed and actioned?</p> <p>How often will your responses to this canvas be reviewed or updated? When?</p>	<p>Your actions</p> <p>What actions will you take before moving forward with this project? Which should take priority?</p> <p>Who will be responsible for these actions, and who must be involved?</p> <p>Will you openly publish your actions and answers to this canvas?</p>



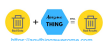
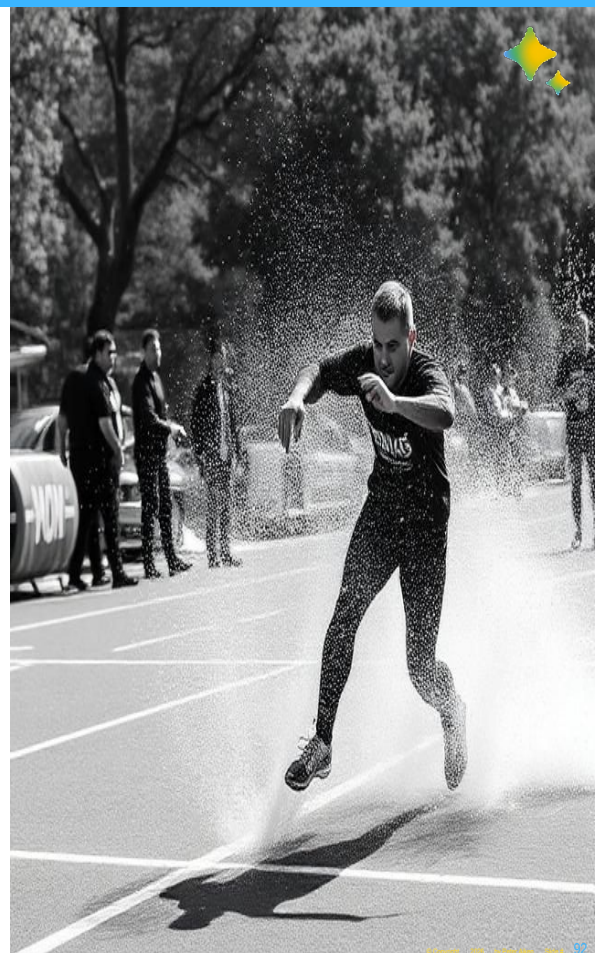
In Action

Question:

- What are any possible ethical considerations about posting some email addresses online at my website

Response:

- While posting these emails might not be "illegal" in the strictest sense, the Data Ethics Canvas suggests it is high-risk.
- If you want to maintain the integrity of the website, I recommend focusing on the actions or roles of the individuals rather than their raw contact data.
- Facilitate communication without facilitating exploitation.



Describe How Data Stewards Can Use AI To Augment Their Roles



As organizations shift from manual oversight to proactive, automated data management, data stewards can use Artificial Intelligence (AI) to transition from "data gatekeepers" to strategic partners who directly support business objectives. By integrating AI, stewards can move beyond manual, reactive tasks to focus on overseeing AI-driven processes and ensuring ethical data use.

1. Automating Data Quality and Validation

AI and machine learning allow stewards to manage data quality at scale by moving from manual, rule-based checks to predictive and automated cleansing.

- **Anomaly Detection:** AI identifies subtle patterns and sudden shifts in data distributions (such as unusual transactions or missing values) faster than human stewards, particularly across massive datasets.
- **Predictive Cleansing:** Machine learning models can automatically clean, de-duplicate, and standardize records (e.g., identifying that "John A. Smith" and "J. Smith" are the same person).
- **Filling Gaps:** AI can intelligently estimate or fill in missing values based on contextual patterns within the rest of the dataset.

2. Enhancing Metadata and Business Glossaries

AI transforms metadata from static documentation into operational intelligence, making data discovery more efficient for the entire organization.

- **Automated Tagging:** AI can infer semantics (e.g., recognizing that a column labeled "cust_id" is a customer identifier) and automatically apply classification tags for sensitivity or compliance.
- **Semantic Matching:** Natural Language Processing (NLP) helps identify synonymous terms across departments, assisting stewards in maintaining a consistent **Business Glossary**.
- **Natural Language Discovery:** Business users can use plain English to search metadata (e.g., "Show me dashboards depending on this Snowflake table"), reducing the reliance on centralized discovery teams.



3. Strengthening Data Ethics and Risk Management

As guardians of an organization's ethical data culture, stewards use AI to identify risks and ensure fairness in automated decision-making.

- **Bias Detection:** AI audits datasets for inherent biases—often stemming from skewed training data—to prevent discriminatory outcomes in areas like recruitment or financial lending.
- **Compliance Monitoring:** AI-powered systems continuously monitor access patterns to confidential data and alert stewards of unauthorized attempts or access from unapproved geographic locations.
- **Automated Policy Enforcement:** AI can automatically apply role-based access controls (RBAC) and propagate sensitivity tags across the data estate based on established policies.

4. Support for Data Science and Lineage

Data stewards provide the reliable foundation upon which data scientists build models by ensuring transparency and traceability.

- **Auto-Constructed Lineage:** AI automatically maps data flows across systems by analyzing queries and logs, creating a "tree of traceability" that shows exactly where data originated and how it was transformed.
- **Impact Analysis:** Stewards use automated lineage graphs to immediately diagnose the root cause of data quality issues or determine which downstream assets will break if a table is changed.
- **Explainable AI (XAI):** Stewards help maintain public trust by using techniques to ensure that algorithms remain transparent.

5. Demonstrating Strategic Value

AI helps stewards quantify the impact of their work in economic terms, shifting their reputation from a cost center to a value provider.

- **Quantifying the "Cost of Bad Data":** AI can model the financial impact of data quality improvements on key business KPIs, such as operational efficiency or customer retention.
- **Scaling Governance:** By automating routine tasks, stewards are freed to focus on high-value strategic initiatives that align data management with the overall corporate vision.

© Copyright 2020 by Peter Allen Slide 4 93

What are some of the first duties of the first data stewards in any organization?



The first duties of the initial data stewards in any organization are crucial for setting the foundation of a successful data governance program. Here's a breakdown of key early responsibilities:

1. Understanding the Data Landscape:

- **Data Inventory & Discovery:** Identify and document the organization's critical data assets. This includes locating where data resides (databases, files, applications), what types of data are collected, and how it's used. This is a foundational step.
- **Data Profiling:** Analyze the quality, completeness, consistency, and accuracy of the data. Understanding the "as-is" state is essential for improvement.
- **Business Process Mapping:** Understand how data flows through key business processes. This helps identify data dependencies and potential bottlenecks.
- **Stakeholder Identification:** Identify key stakeholders who create, use, and are impacted by the data. Understanding their needs and perspectives is crucial.

2. Defining Data Governance Framework Elements:

- **Data Governance Policies & Standards (Initial Drafts):** Collaborate with stakeholders to develop initial drafts of data governance policies and standards. These will likely evolve over time but starting with core principles is key. Focus on what is most important initially.
- **Data Definitions & Business Glossary (Initial Version):** Begin building a business glossary to define key data terms and ensure everyone is on the same page. This is a living document that will grow.
- **Data Ownership & Stewardship Assignment (Initial):** Work with leadership to assign initial data ownership and stewardship responsibilities. This may need refinement as the organization's data governance maturity increases.
- **Data Quality Rules (Initial Set):** Define a starting set of data quality rules to address the most pressing data quality issues. These rules should be measurable and actionable.

3. Building Relationships and Communication:

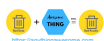
- **Stakeholder Engagement:** Communicate the importance of data governance to key stakeholders and build support for the program. Early wins and visible progress are crucial here.
- **Collaboration with IT:** Establish a strong working relationship with the IT department, as they are essential for implementing data governance solutions.
- **Data User Training (Basic):** Provide basic training to data users on data governance policies and procedures. Focus on what is immediately relevant.

4. Demonstrating Value and Iteration:

- **Quick Wins:** Focus on achieving some quick wins to demonstrate the value of data governance. Address a few key data quality issues or improve a critical data-dependent process. Visibility is key.
- **Establish Metrics:** Define metrics to track the effectiveness of data governance efforts. This will help demonstrate ROI and justify continued investment.
- **Iterative Approach:** Recognize that data governance is a journey, not a destination. Be prepared to iterate and refine the framework based on feedback and experience.

Key Considerations for Initial Data Stewards:

- Focus on the most critical data first. Don't try to boil the ocean.
- Prioritize collaboration and communication. Building relationships is essential.
- Start small and iterate. Don't try to implement everything at once.
- Demonstrate value early and often. This will help build support for the program.
- Be patient and persistent. Data governance takes time and effort.
- By focusing on these initial duties, data stewards can lay a solid foundation for a successful data governance program and help the organization realize the full potential of its data assets.



How Can Data Stewards Use AI To Value Data Challenges?



While the DAMA-DMBOK framework primarily focuses on foundational data management practices, it highlights how data stewards can leverage advanced technologies like AI to address critical "Data Challenges" and enhance organizational value. By integrating AI into stewardship activities, you can move from manual oversight to proactive, automated management that directly supports high-level business objectives.

Leveraging AI for Core Data Challenges

According to the framework, AI can be applied to several key areas where data stewards provide value:

- **Automated Data Quality Management:** Instead of manual cleansing, stewards can use AI-driven tools for anomaly detection and predictive cleansing. AI can automatically flag outliers or suspicious data patterns that standard business rules might miss, continuously profiling data "health" in real-time.
- **Intelligent Metadata and Business Glossaries:** AI can assist in the automated tagging and classification of unstructured data (e.g., documents and emails). Large language models can also help align diverse vocabularies through semantic matching, making it easier to maintain a consistent Business Glossary across different departments.



- **Mitigating Ethical Risks and Bias:** Stewards act as guardians of ethical data culture. AI tools can be used for bias detection in existing algorithms and datasets, ensuring fair automated decisions. AI-powered discovery can also find hidden sensitive data (PII) across the enterprise to ensure regulatory compliance with laws like GDPR.
- **Support for Advanced Analytics and Data Science:** Data stewards provide the reliable foundation for AI models. By using AI to document data lineage, stewards provide the transparency needed to trust "black box" models. They also support feature engineering by guiding AI teams toward the most relevant business data attributes.

Transitioning to Strategic Value

To be seen as a provider of value, data stewards should use AI to shift their focus from reactive tasks to strategic partnerships.

- **Value Quantification:** AI can help stewards model the "cost of bad data" (e.g., wasted spend or compliance fines), providing a stronger business case for stewardship investments.
- **Predictive Stewardship:** By adopting AI-driven tools, stewards transition from reactive issue management to proactive, predictive activities that improve operational efficiency and support innovation.

Advice for Mentees: Always align AI initiatives with specific business pain points. AI should not be implemented for its own sake, but as a tool to make organizational data more accessible, reliable, and ethically sound—thereby establishing your role as a vital contributor to organizational success.

Specific Data Steward ↔ AI Integrations

- Complement architectural (re)developmental opportunities
- Understanding the cadence/role of foundational data practices supporting organizational operations
- Proving value with tangible ROI
- Improving AI effectiveness/efficiencies using organization-wide perspectives
- Comprehending how stewards need to be multifunctional and dexterous, especially at first
- Integrating the role of data debt fighting



Program Overview



- AI Reality Check?
 - Barriers to entry
 - Not “fit for purpose”
 - Data debt
- Specifics (part of a larger effort)
 - Compliment architectural (re)developmental opportunities
 - Understanding the cadence/role of foundational data practices supporting organizational operations
 - Proving value with tangible ROI
 - Improving AI effectiveness/efficiencies using organization-wide perspectives
 - Comprehending how stewards need to be multifunctional and dexterous, especially at first
 - Integrating the role of data debt fighting
- Practical Uses
 - Vision/approach cohesion
 - Build Multifunctional Dexterity
 - Ethical support
- Take aways → Q&A

**Data Stewards:
Conquering Data Debt
in the Age of AI**

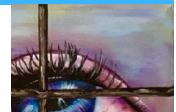


<https://creativecommons.org/licenses/by-nc-sa/4.0/>



© Copyright 2020 by Peter Allen Slide # 97

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



External Comprehension

Everything Else Data

Data Management

Data Governance Program

Data Stewards

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://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2020 by Peter Allen Slide # 98

Why is communication and external comprehension so important to successful implementation of data stewards?



Communication and external comprehension are absolutely vital to the success of data stewards for several interconnected reasons:

1. Building Buy-In and Support:

- Data governance is not just a technical exercise. It's a cultural shift that requires buy-in from all stakeholders. Data stewards need to effectively communicate the why behind data governance – how it benefits the organization, improves decision-making, reduces risks, etc.
- Explaining the value proposition is crucial. People are more likely to support initiatives they understand. Data stewards must articulate the tangible benefits of good data governance in terms that resonate with different audiences (e.g., improved efficiency, cost savings, better customer service).
- Addressing concerns and resistance proactively. Change can be unsettling. Data stewards need to be able to address concerns, explain how the changes will impact different teams, and alleviate any resistance to new processes.

2. Facilitating Collaboration and Coordination:

- Data governance is a team sport. Data stewards need to work with various stakeholders across the organization, including IT, business units, legal, and compliance. Effective communication is essential for coordinating efforts and ensuring everyone is aligned.
- Breaking down silos. Data often resides in different departments and systems. Data stewards need to be able to bridge these silos, fostering collaboration and information sharing across the organization.
- Building consensus. Data governance decisions often require input from multiple stakeholders. Data stewards need to facilitate discussions, gather feedback, and build consensus around key issues.

3. Ensuring Understanding and Compliance:

- Data governance policies and procedures are only effective if people understand them. Data stewards need to communicate these policies clearly and concisely, using language that is accessible to all stakeholders.

- Training and education are crucial. Data stewards need to develop and deliver training programs to ensure data users understand their roles and responsibilities in data governance.

- Reinforcing best practices. Communication is ongoing. Data stewards need to regularly communicate updates, reminders, and best practices to keep data governance top of mind.

4. Driving Data Literacy:

- Data is only valuable if people know how to use it effectively. Data stewards play a key role in promoting data literacy within the organization. They can do this by sharing knowledge, providing training, and advocating for data-driven decision-making.
- Empowering data users. By improving data literacy, data stewards empower data users to make better decisions, solve problems more effectively, and contribute to the organization's success.

5. Demonstrating Value and ROI:

- Data governance initiatives need to demonstrate their value to the organization. Data stewards need to communicate the positive impact of data governance efforts, such as improved data quality, reduced costs, and better decision-making.
- Celebrating successes. Publicly acknowledging successes, even small ones, helps build momentum and reinforces the importance of data governance.

In summary, communication and external comprehension are not just "nice-to-haves" for data stewards – they are fundamental to their success. Without effective communication, data governance initiatives are likely to fail, regardless of how technically sound they may be. Data stewards are essentially change agents, and effective communication is the cornerstone of successful change management.

Data (blah blah blah)



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2020 by Peter Allen Slide 99

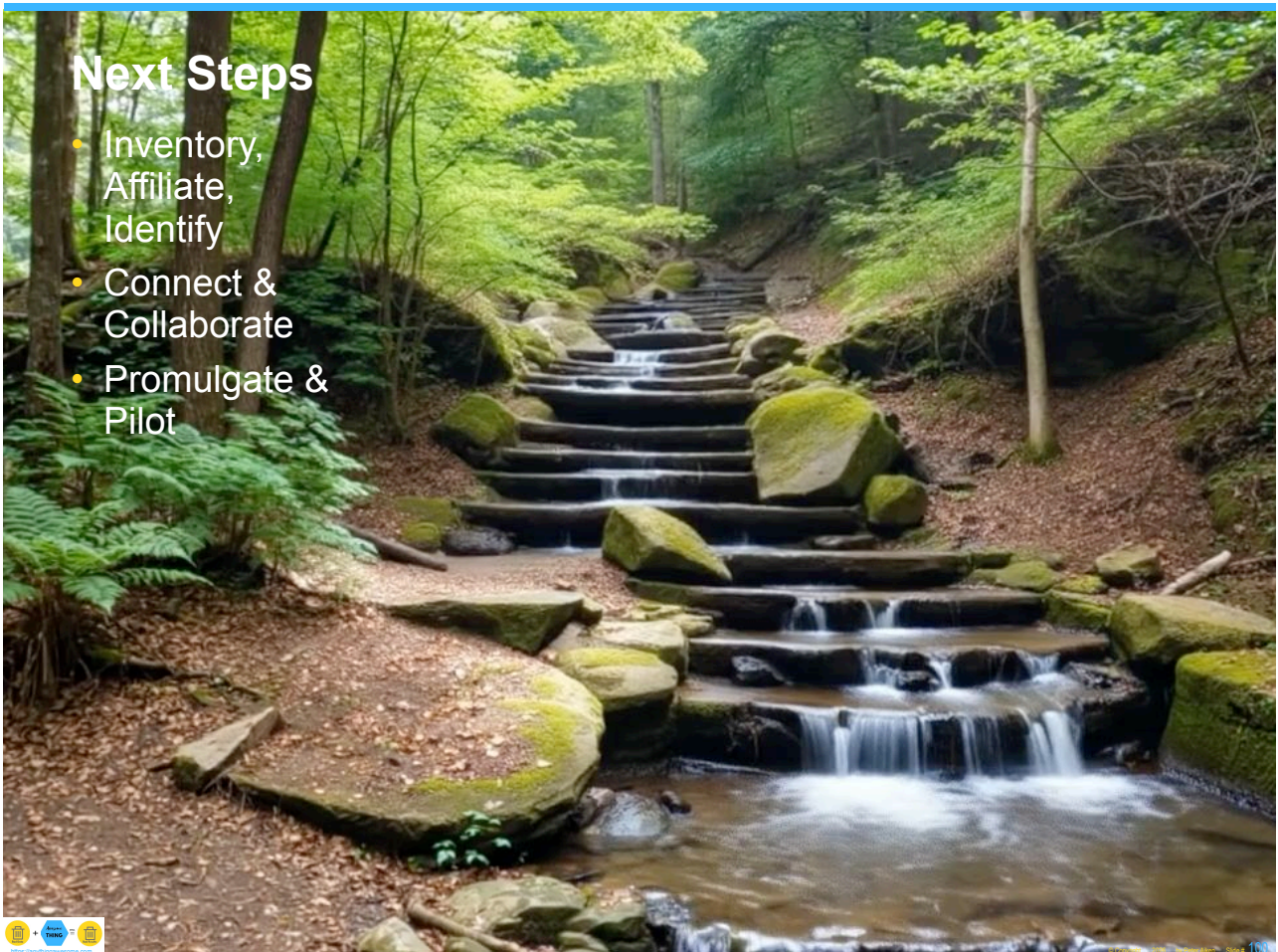
Next Steps

- Inventory, Affiliate, Identify
- Connect & Collaborate
- Promulgate & Pilot



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2020 by Peter Allen Slide 100



Upcoming Events

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

The Importance of Ethics in Data/AI Governance 12 May 2026

Data Integration Demystified: Enabling AI-Driven Interoperability 9 June 2026



AI-Powered Data Modeling Essentials 14 July 2026

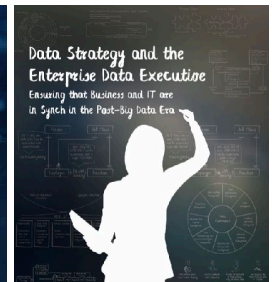
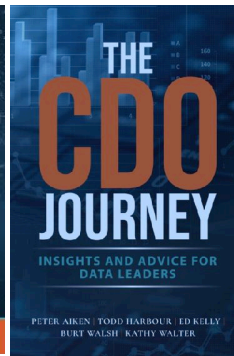
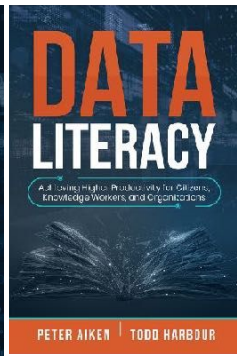
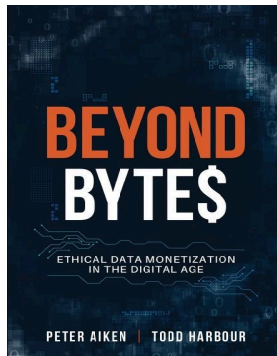
Brought to you by:

[Clicking any webinar title will link directly to the registration page]

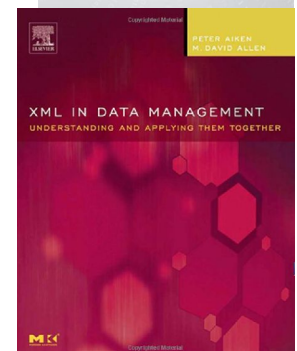
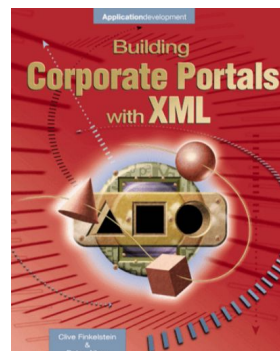
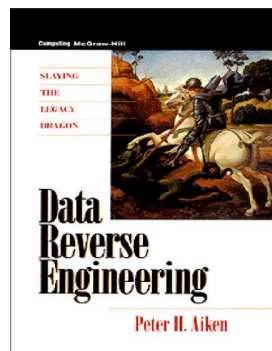
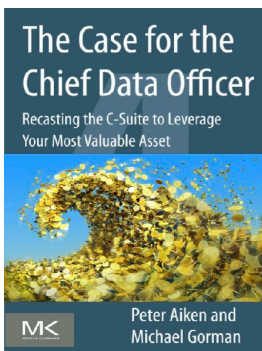
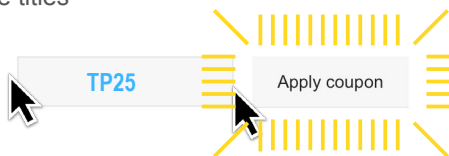


© Copyright 2026 by Peter Aiken. Slide 101

Event Pricing on Peter's Books



- 25% off directly from the publisher on some titles
- My 'Book Store' @ <https://anythingawesome.com/books>
- Enter the code "TP25" at the Technics bookstore checkout where it says to "Apply Coupon"



Independent Verification & Validation



Critical Design Review?

Mentoring?

Executive/Data Literacy Training?

Collaboration?



+



STILL =



Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Data Governance Training?

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>

