The Roles of Data Stewards

Architects of the Data-Driven Future

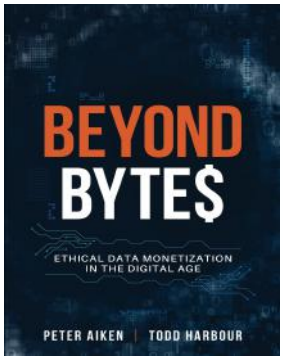
peter.aiken@anythingawesome.com +1.804.382.5957

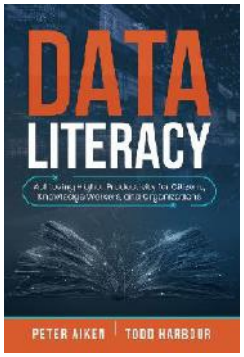
Peter Aiken, Ph.D.

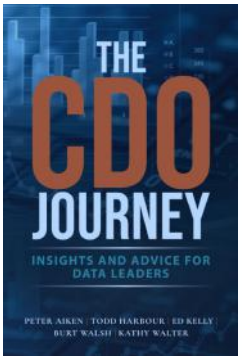
- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices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

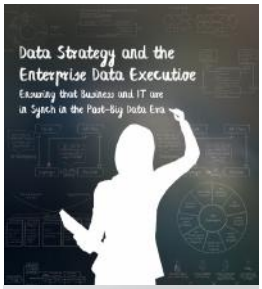
$1,500,000,000.00 USD

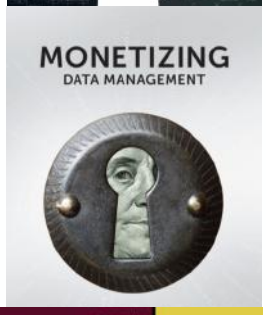
Event Pricing on Peter's Books



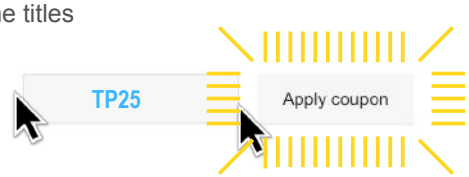


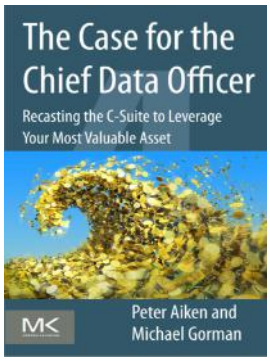


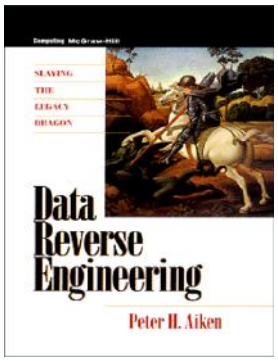


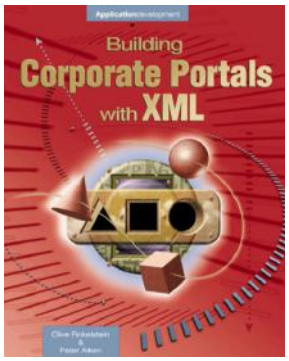


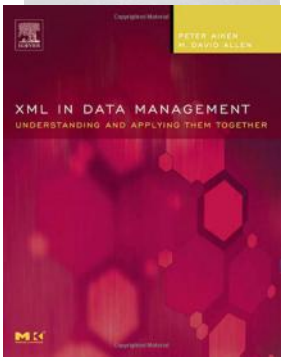
- 25% off directly from the publisher on some titles
- My 'Book Store' @
  https://anythingawesome.com/books
- Enter the code
  "TP25" at the Technics bookstore
  checkout where it says to "Apply Coupon"











Program Overview

- Why do we need data stewardship as a role?
  - Definitions: Stewardship, data stewards, data debt
  - The role of strategy
  - Data architecture = strategic focus implemented by stewards
- What are they supposed to do?
  - Resolve prerequisite challenges stemming from data debt
  - Data stewardship framework
  - Fire station model (Reactive & proactive foci)
  - Stewardship role (in context of data governance)
- Assigning data stewards (tangible improvement)
  - Start simply
  - Differing cadence (Need for different structural approach)
  - Foundational prerequisites
  - Need for simplicity, agility, practice
- Take aways ➜ Q&A

The Roles of Data Stewards

Question?

STARTING OR RESTARTING

- How many starting versus how many re-starting?

Definitions

Steward

1. a person who looks after the passengers on a ship, aircraft, or train and brings them meals.
   - synonyms: flight attendant, cabin attendant, air hostess, purser "an air steward"
   - a person responsible for supplies of food to a college, club, or other institution
2. an official appointed to supervise arrangements or keep order at a large public event, for ex. sporting event.
   - synonyms: official, marshal, organizer "the race stewards"
   - short for shop steward
3. **a person employed to manage another's property**, especially a large house or estate.
   - synonyms: (estate) manager, agent, overseer, **custodian**, **caretaker**; historical "**the steward of the estate**"
   - a person whose responsibility it is to take care of something."farmers pride themselves on being **stewards** of the countryside"

Stewarding

1. (of an official) supervise arrangements or keep order at (a large public event).
   "the event was organized and stewarded properly"
2. **manage or look after** another's property).

Data Steward

- Manage data assets on behalf of all stake holders and in the best interests of the organization (McGilvray, 2008)
- Represent the interests of all stakeholders and take an enterprise perspective
- Have dedicated time enough to be accountable and responsible

Trust

- Firm belief in the reliability, truth, ability, or strength of someone or something (google.com)

Fiduciary

- Involving trust, especially with regard to the relationship between a trustee and a beneficiary (google.com)

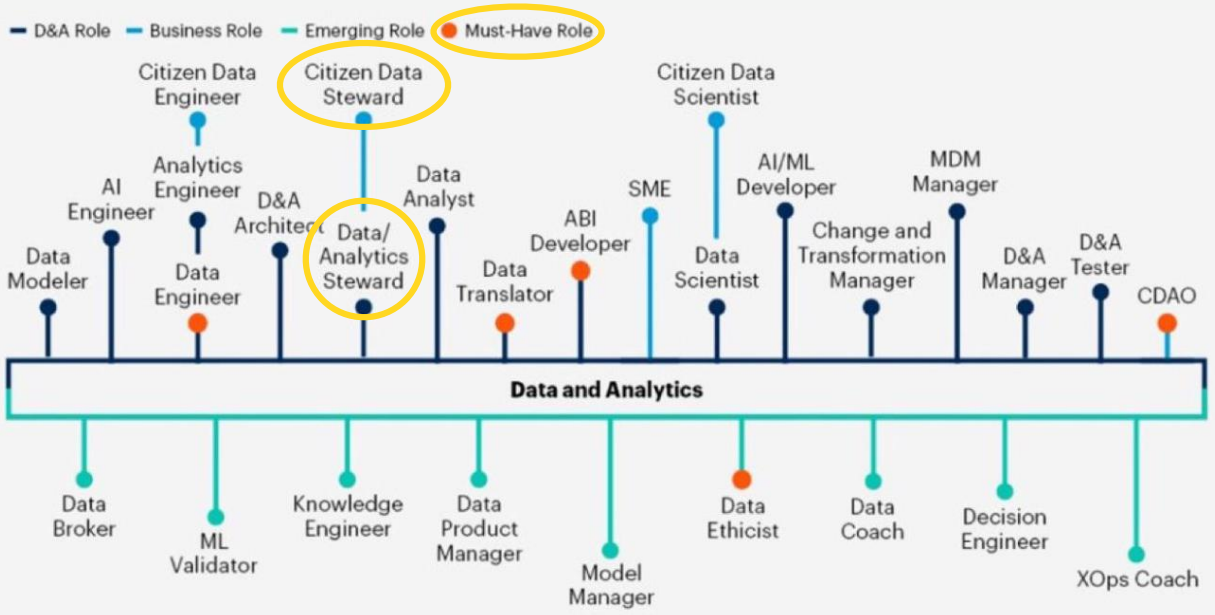
Steward, Data

- one who actively directs the use of organizational data assets in support of specific mission objectives

Merriam-Webster

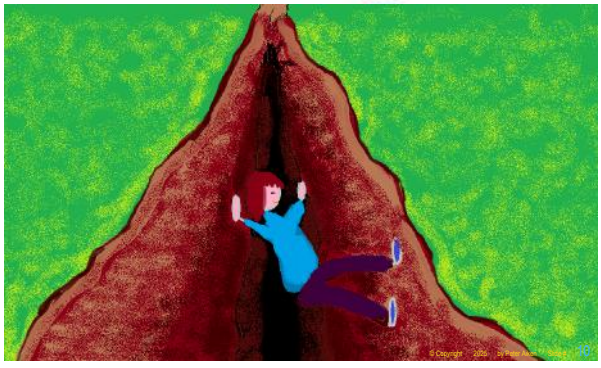
INTERNATIONAL DATA STEWARDS DAY

The emerging spectrum of data and analytics roles



Confusion as to data responsibility

- IT thinks data is a business problem
  - *"If they can connect to the server, then my job is done!"*
- The business thinks IT is managing data adequately
  - *"Who else would be taking care of it?"*



Reference instead the term: fiduciary relationship

(Speak of owning data requirements)

What data would accounting own?

Fiduciary Relationship

- What is a Fiduciary?
  - A fiduciary is a person who is legally obligated to act in the best interest of the individual, group, or company. This includes:
  - Lawyers
  - Trustees
  - Doctors
  - Accountants
  - Corporate directors
  - Legal guardians
- Three top duties
  - The Duty to Act in Good Faith
  - The Duty of Care
  - The Duty of Loyalty
- Fiduciaries have legal and ethical standards that they must adhere to

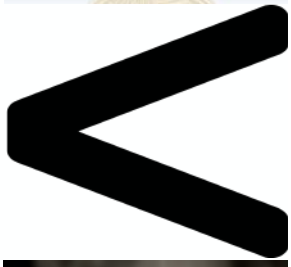
https://mint.intuit.com/blog/planning/what-is-a-fiduciary/

The Princess on the Pea

by
Hans Christian Andersen

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)



Forbes

American Airlines

UNITED AIRLINES

- $6b market value
- $30b AAdvantage
- $9b market value
- $22b MileagePlus

EDITORS' PICK | Jul 15, 2020, 09:00am EDT

How Airlines Make Billions From Monetizing Frequent Flyer Programs

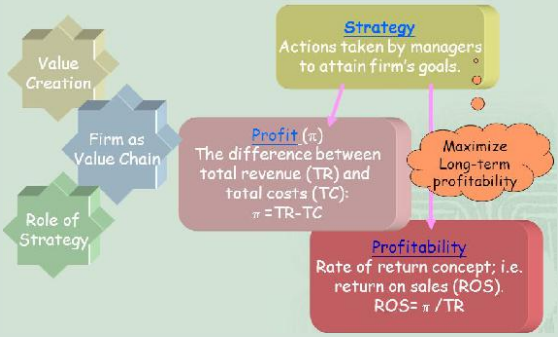
Forbes Advisor

JT Genter Advisor Cont
Advisor Contributor Gro
Personal Finance

How to unlock $24b and $13b in data value?

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

Steward's Role in Data Strategy



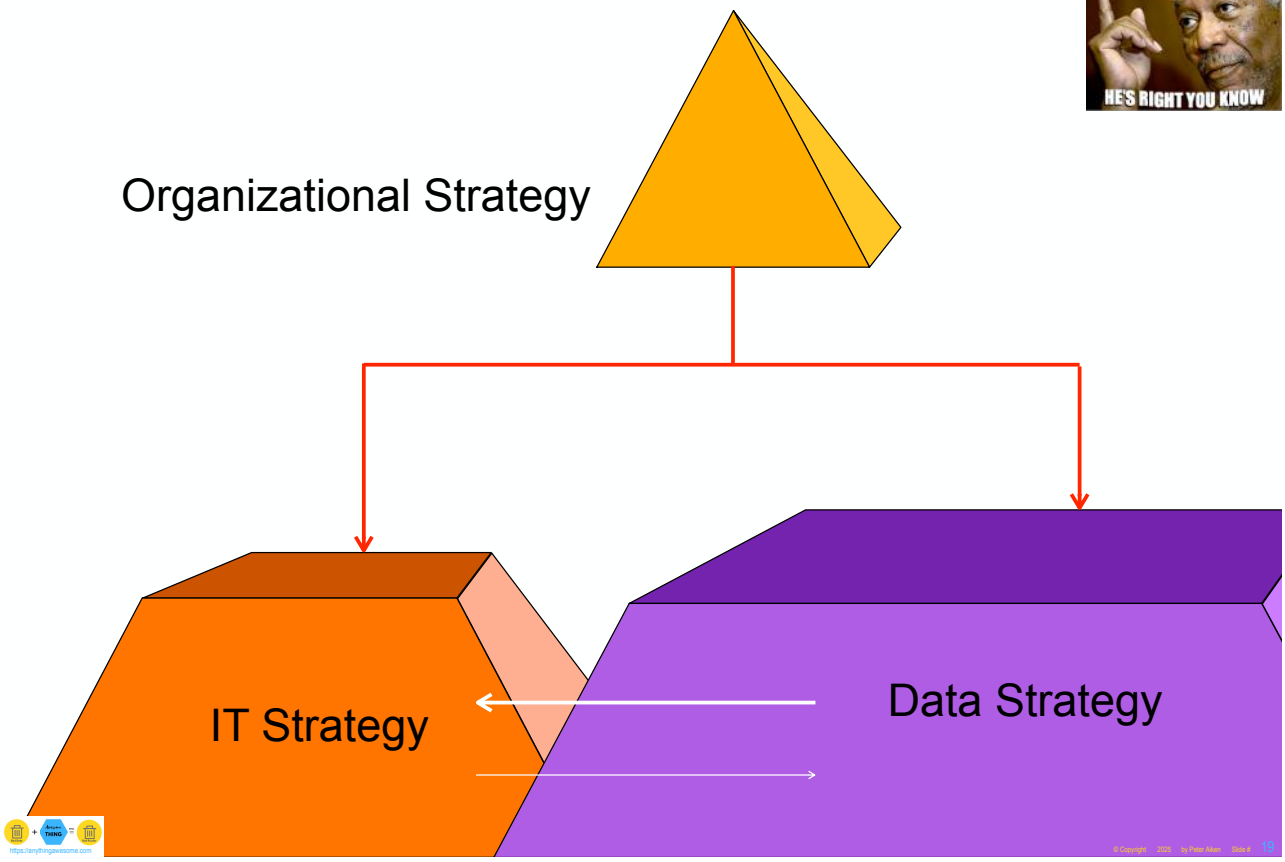
Data Strategy in Context – THIS IS WRONG!

Organizational Strategy

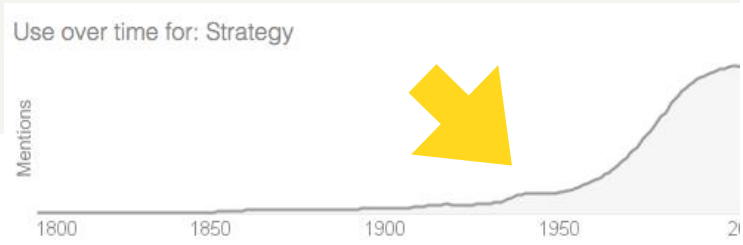
IT Strategy

Data Strategy

This is correct …



What is Strategy?

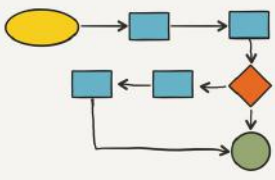


strat·e·gy
/ˈstratəjē/

noun

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

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

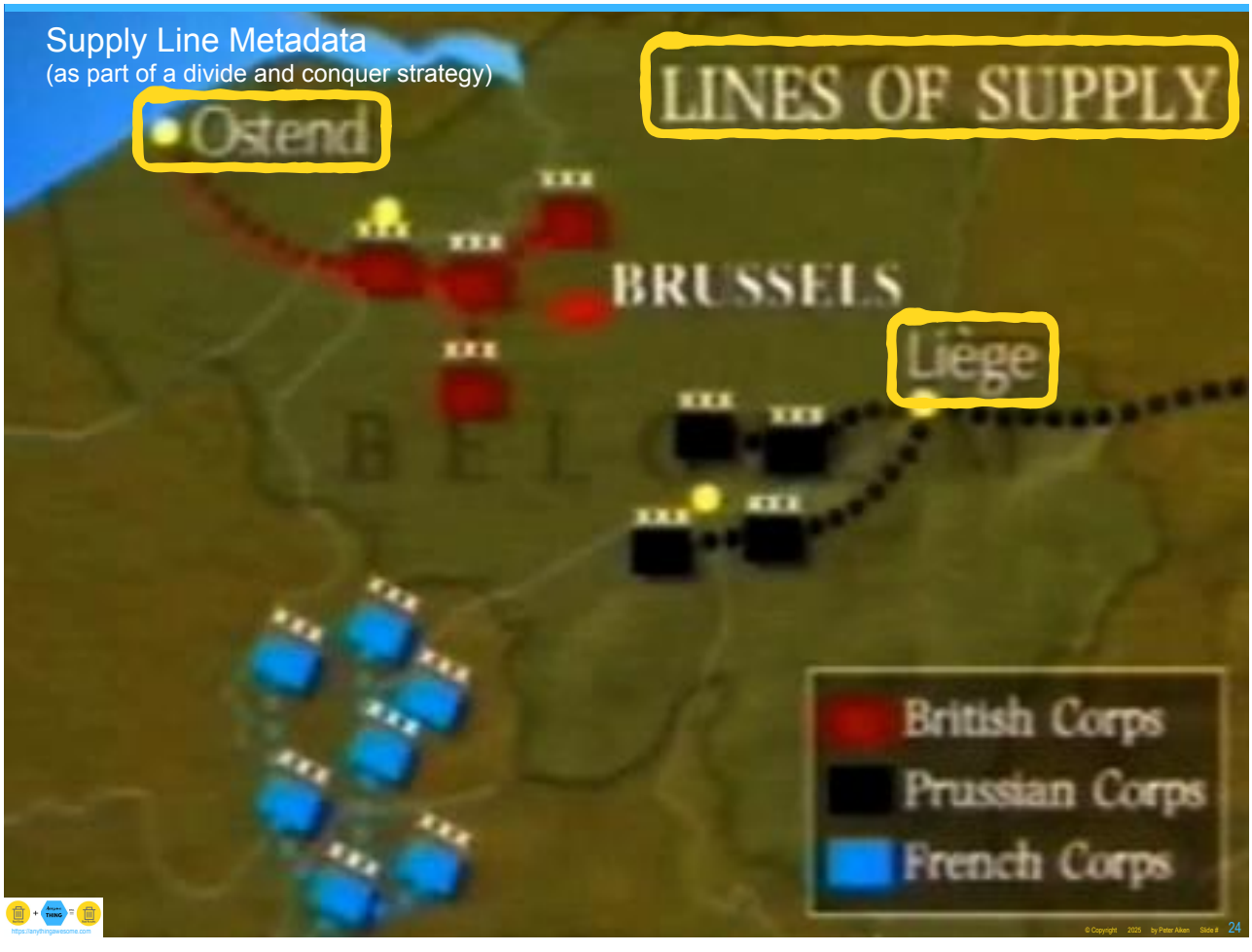


Former Walmart Business Strategy

Every Day Low Price

Strategy in Action: Napoleon faces a larger enemy

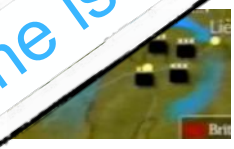
- Question?
  - How do I defeat the competition when their forces are bigger than mine?
- Answer:
  - Divide and conquer!
  - **"a pattern in a stream of decisions"**



Complex Strategy

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

- Then
  - Turn right and the Prussia



- The
  - eft and defeat British

While someone is shooting at you!

Complex Data Governance Environment

DATA COMMISSION

- EXECUTIVE
- Region representatives
- LEGISLATIVE
- JUDICIAL

Advise Governor

Data sharing and analytics
- Identify goals and objectives
- Prioritize initiatives
- Study & report
- Recommend changes to budget and code

TRUSTEE

execute

- Define, approve, and communicate data strategies, policies, standards, rules, guidelines, & best practices
- Provide a governance, policy, and technology framework
- Define agency data governance responsibilities
- Encourage & facilitate data sharing
- Facilitate coordination to prevent duplication
- Coordinate policy and technology proposals and recommendations
- Administer and manage the commonwealth data trust
- Track and enforce compliance and conformance
- Oversee dissemination of open data

Data Governance Council

- Liaise between agency operations & CDO
- Advise CDO on technology, policy, and governance strategies
- Administer data governance policies set by the board
- implement data sharing & analytics projects
- Review open data assets
- Report progress & compliance to the Board

Advise CDO

Executive data board

- Translate commonwealth goals to agency performance targets
- provide resources
- Remove organizational obstacles
- appoint data governance council members
- Oversee the data governance council
- Oversee data sharing & analytics projects

Oversee council

A pattern in a stream of decisions

Governance and Architecture

Example from: https://www.slideshare.net/AnthonyDehnashi/architecture-governance

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://www.forbes.com/sites/johnmacintosh/2019/10/03/jamie-dimon-and-other-ceos-say-shareholder-value-no-longer-top-priority-epiphany-or-pr-stunt/?sh=17d2895a5497

Maximizing shareholder value can no longer be a company’s main purpose: top CEOs

Published: Aug 24, 2019 1:00 p.m. ET

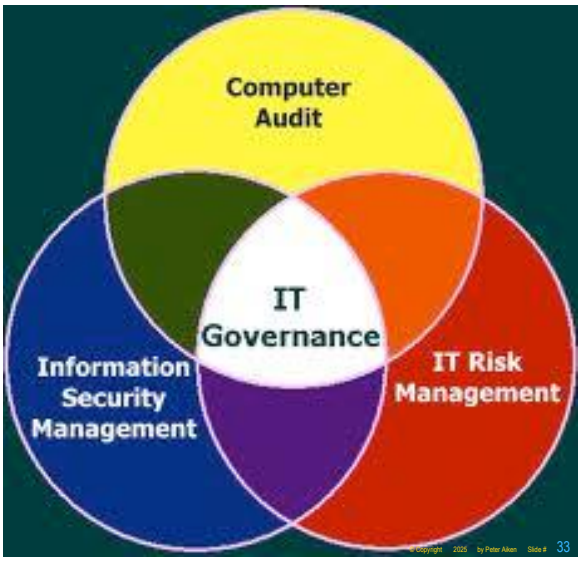
CEOs split from a Milton Friedman mantra, expanding their view of all stakeholders that matter — including workers

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**

Elevator Pitch

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

What is Data Governance?

Managing Data with Guidance

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

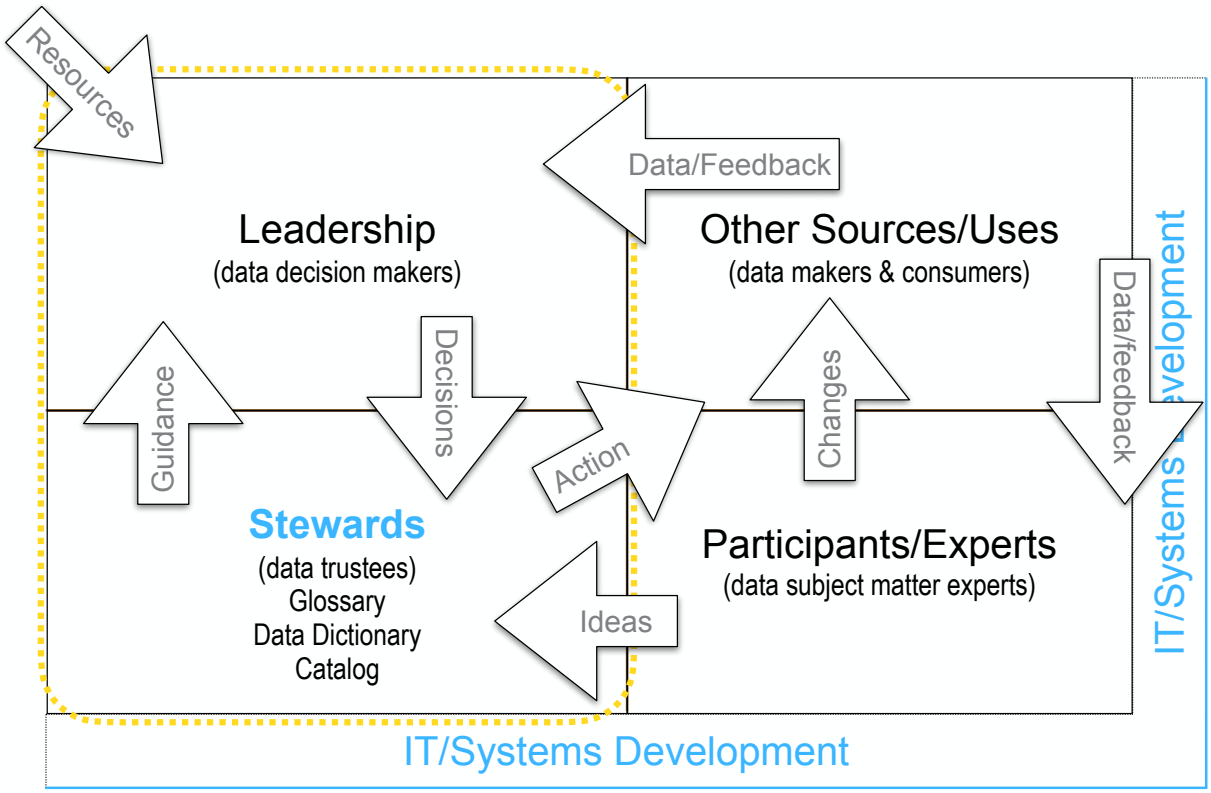
What is Data Governance?

Managing Data Decisions with Guidance

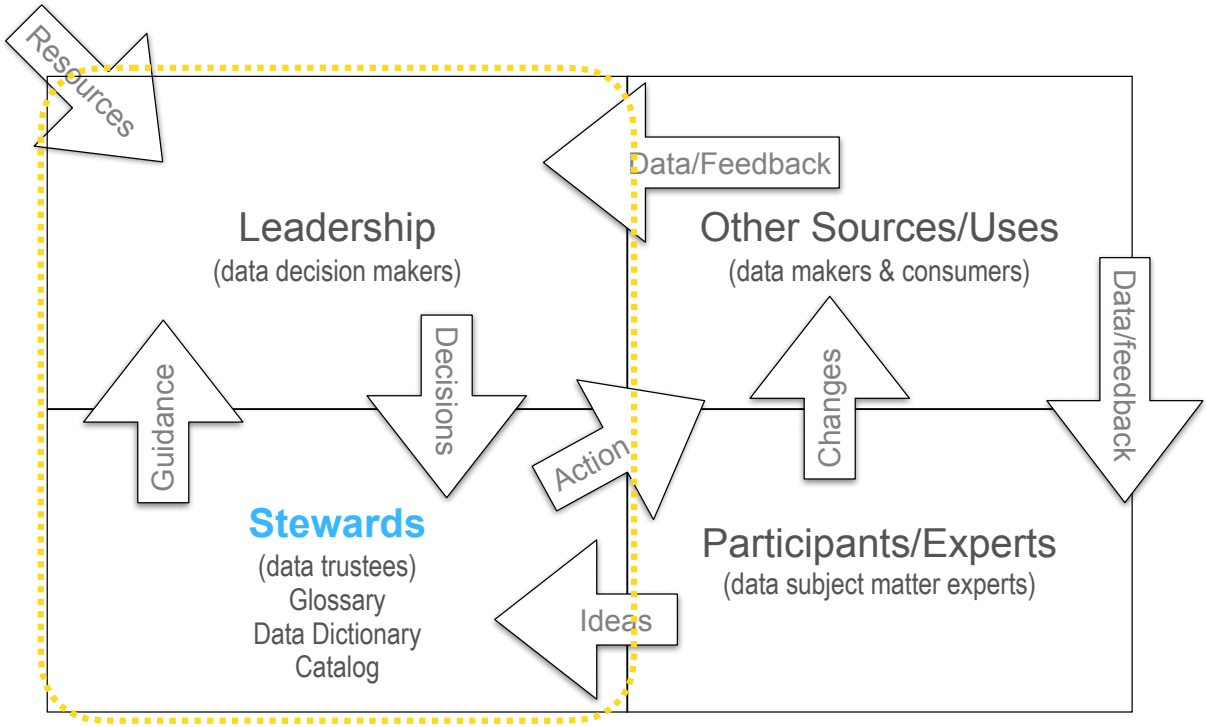
Go Ask Anyone!™

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

Data architecture = strategic focus implemented by stewards



Data architecture = strategic focus implemented by stewards

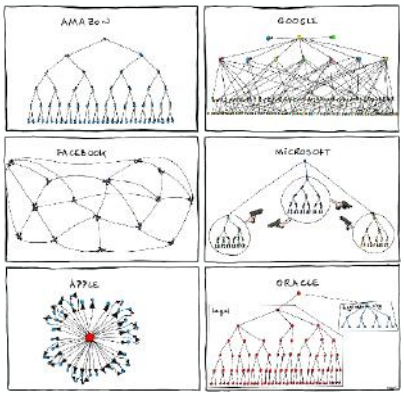


Architecture

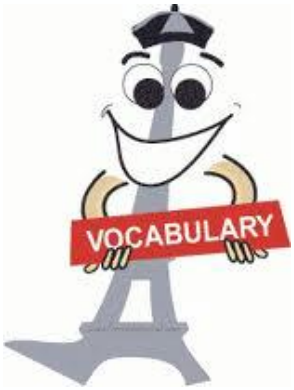
- Things
  - (components)
    data structures
- The functions of the things
  - (individually)
    sources and uses of data
- How the things interact
  - (as a system, towards a goal)
    Efficiencies/effectiveness

Data/Information Architectures – Useful Definition

- Common vocabulary expressing integrated requirements ensuring that data assets are stored, arranged, managed, and used in systems in support of organizational strategy



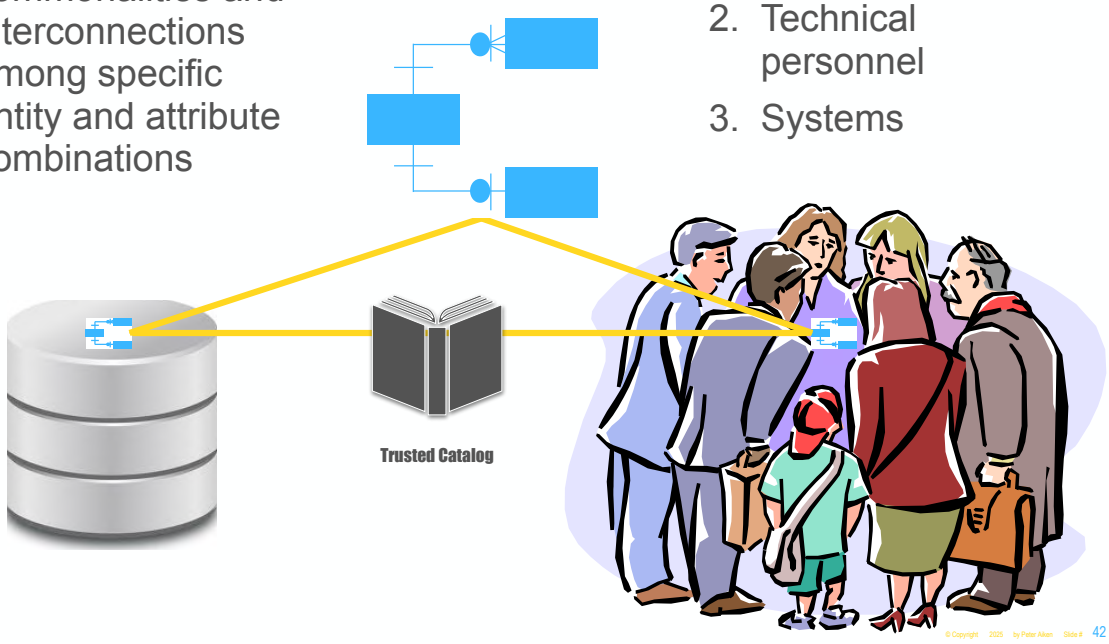
- A primary form of collaboration among data governance professionals and between DG professionals and technical and business professionals



Understanding = Interoperability

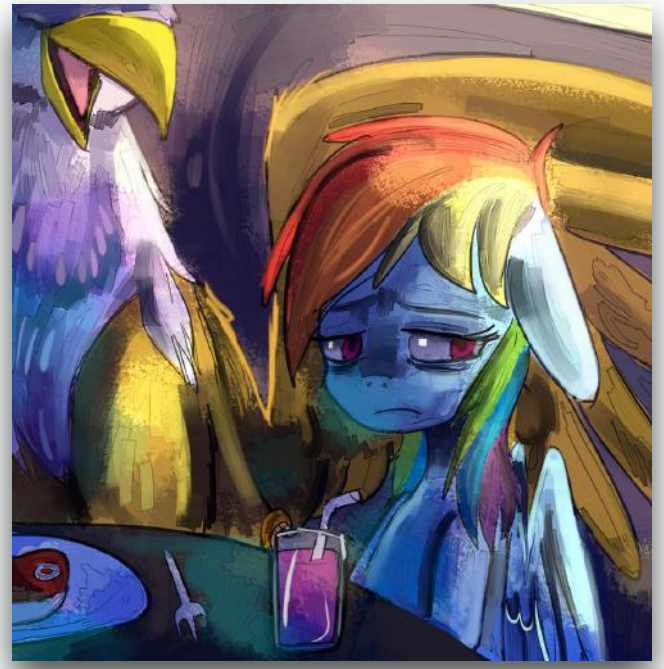
- Documented and articulated as a digital blueprint of the commonalities and interconnections among specific entity and attribute combinations

- Shared by
  1. Business users
  2. Technical personnel
  3. Systems



Architectures: here, whether you like it or not

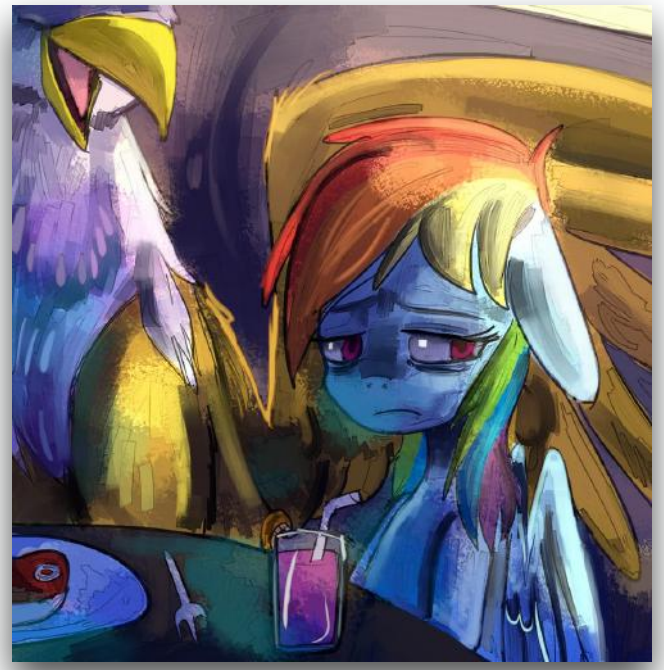
- All organizations have architectures
  - Some are better **understood** and **documented** (and therefore more **useful** to the organization) than others



deviantart.com

Data Architectures: here, whether you like it or not

- All organizations have data architectures
  - Some are better **understood** and **documented** (and therefore more **useful** to the organization) than others



deviantart.com

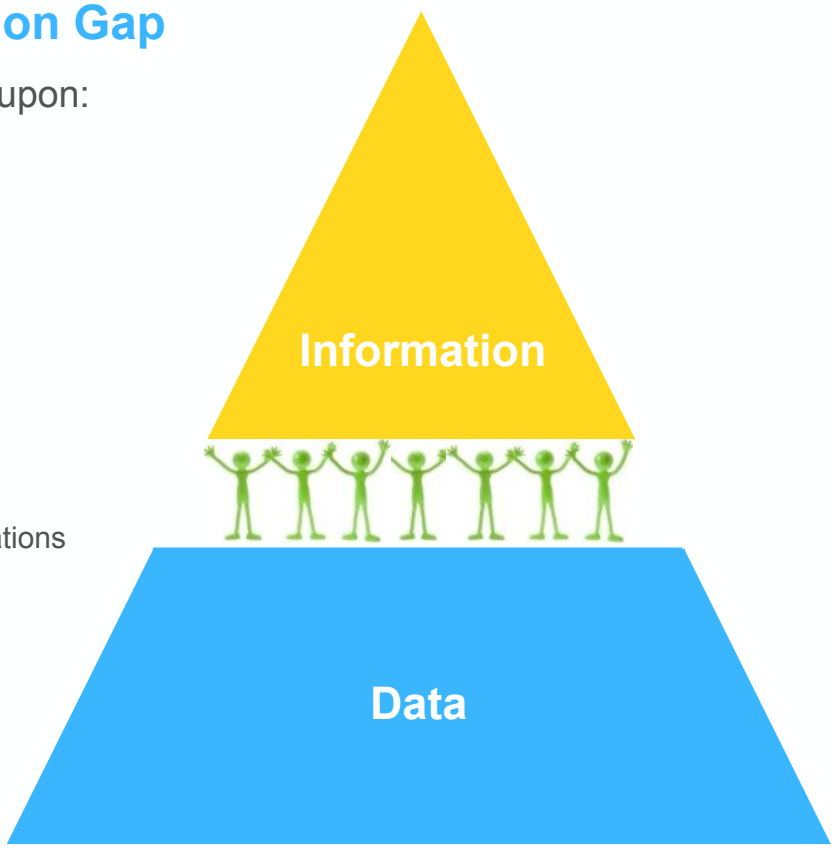
Program Overview

- Why do we need data stewardship as a role?
  - Definitions: Stewardship, data stewards, data debt
  - The role of strategy
  - Data architecture = strategic focus implemented by stewards
- What are they supposed to do?
  - Resolve prerequisite challenges stemming from data debt
  - Data stewardship framework
  - Fire station model (Reactive & proactive foci)
  - Stewardship role (in context of data governance)
- Assigning data stewards (tangible improvement)
  - Start simply
  - Differing cadence (Need for different structural approach)
  - Foundational prerequisites
  - Need for simplicity, agility, practice
- Take aways ➜ Q&A

The Roles of Data Stewards

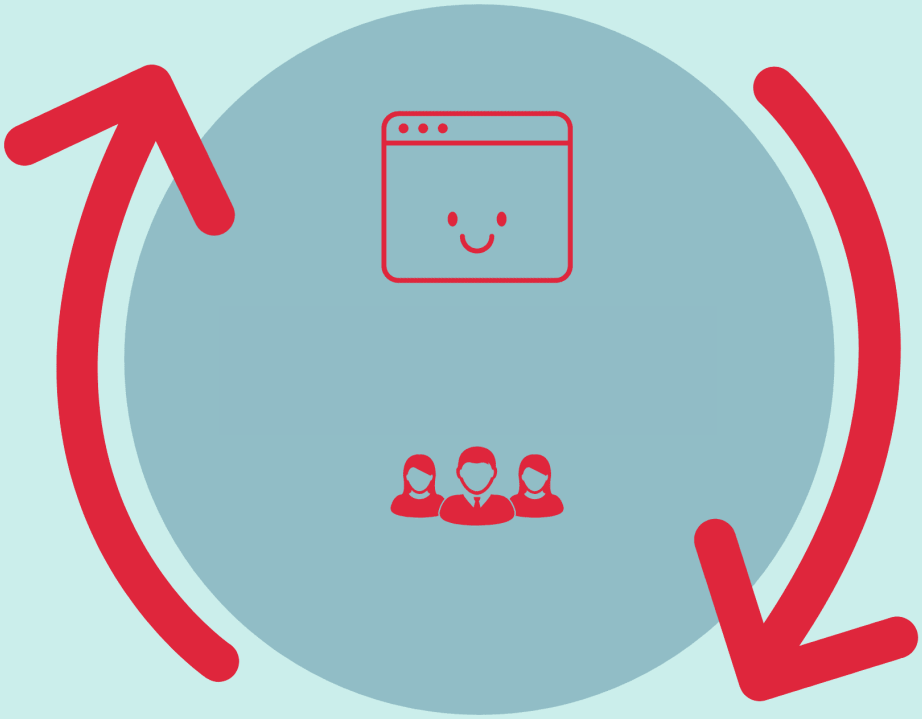
Data / Information Gap

- Overly dependent upon:
  - Human-beings
  - Wetwear
  - Knowledge workers
  - Informal communications
  - Often described as the weakest link



(Especially at first) stewards require a united purpose

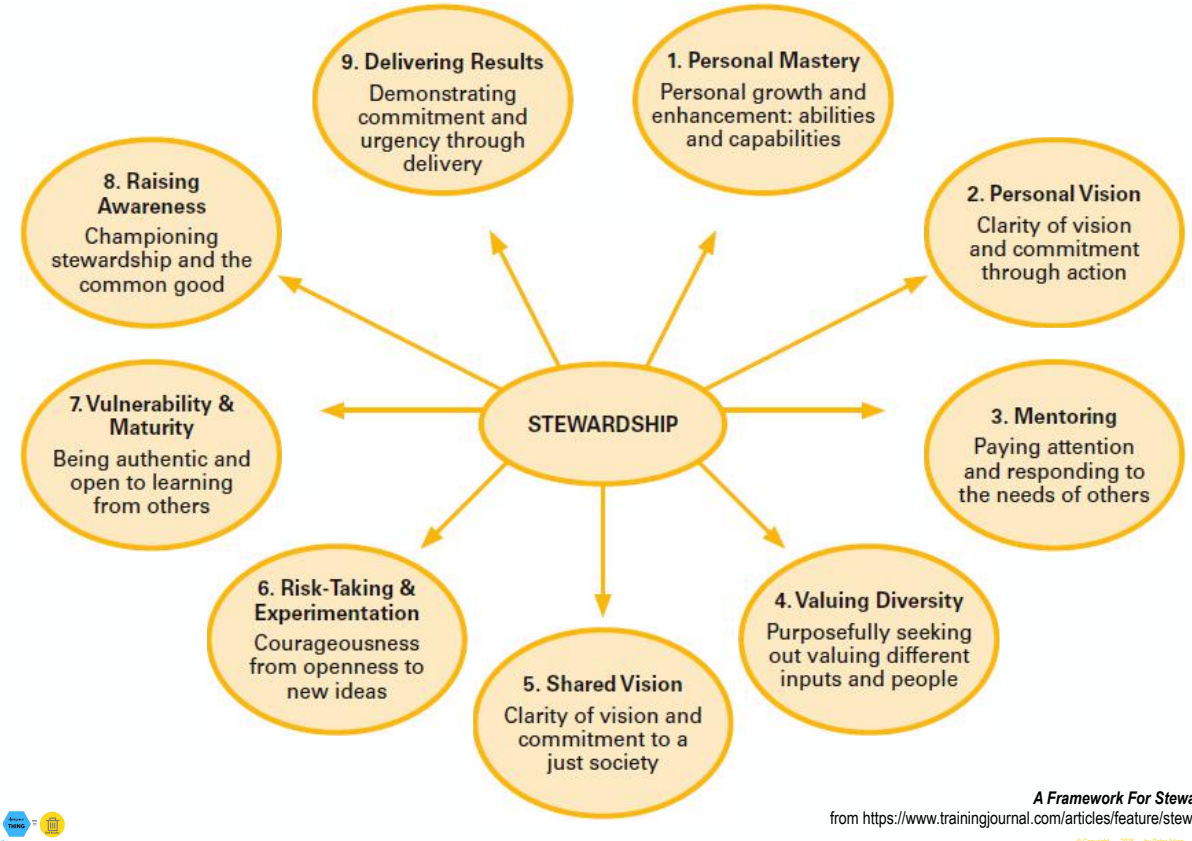
- Enable the organization to better use its data in support of the mission



Data Governance Frameworks

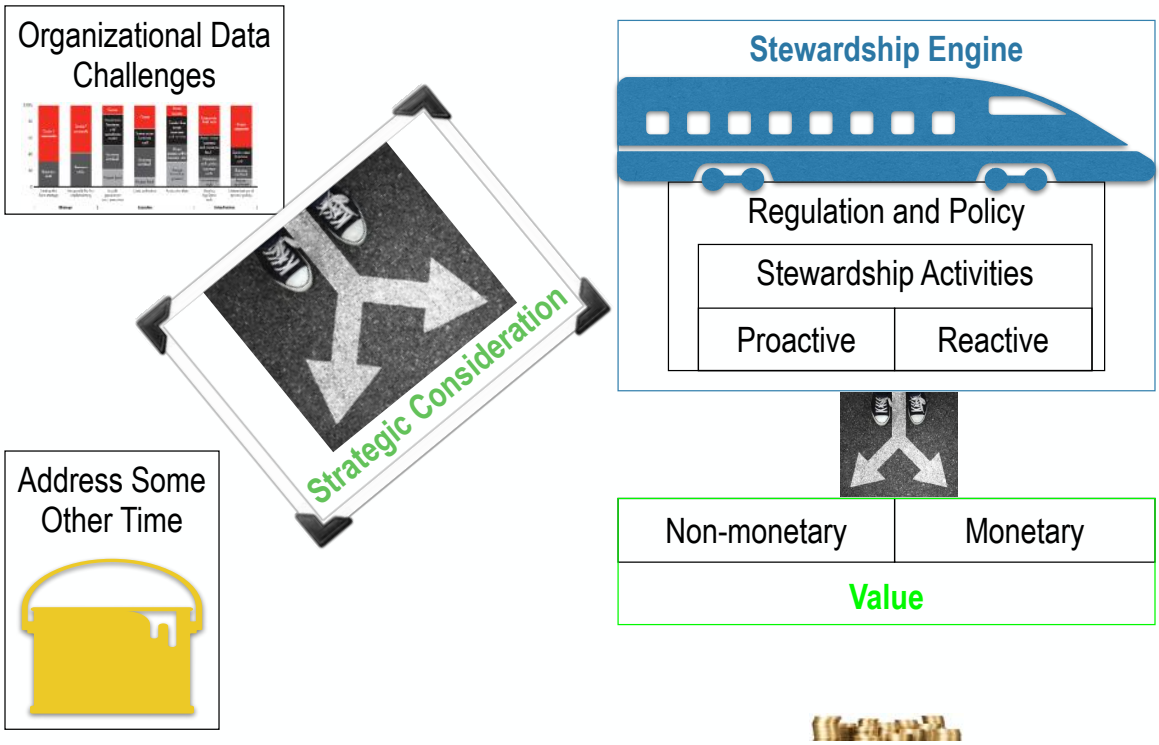
- A system of ideas for guiding analyses
- A means of organizing project data
- Priorities for data decision making
- A means of assessing progress
  - Don't put up walls until foundation inspection is passed
  - Put the roof on ASAP
- Make it all dependent upon continued funding

A Framework For Stewardship



A Framework For Stewardship
from https://www.trainingjournal.com/articles/feature/stewardship

A Framework for Data Stewardship



gemini.google.com

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 (e.g., 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.

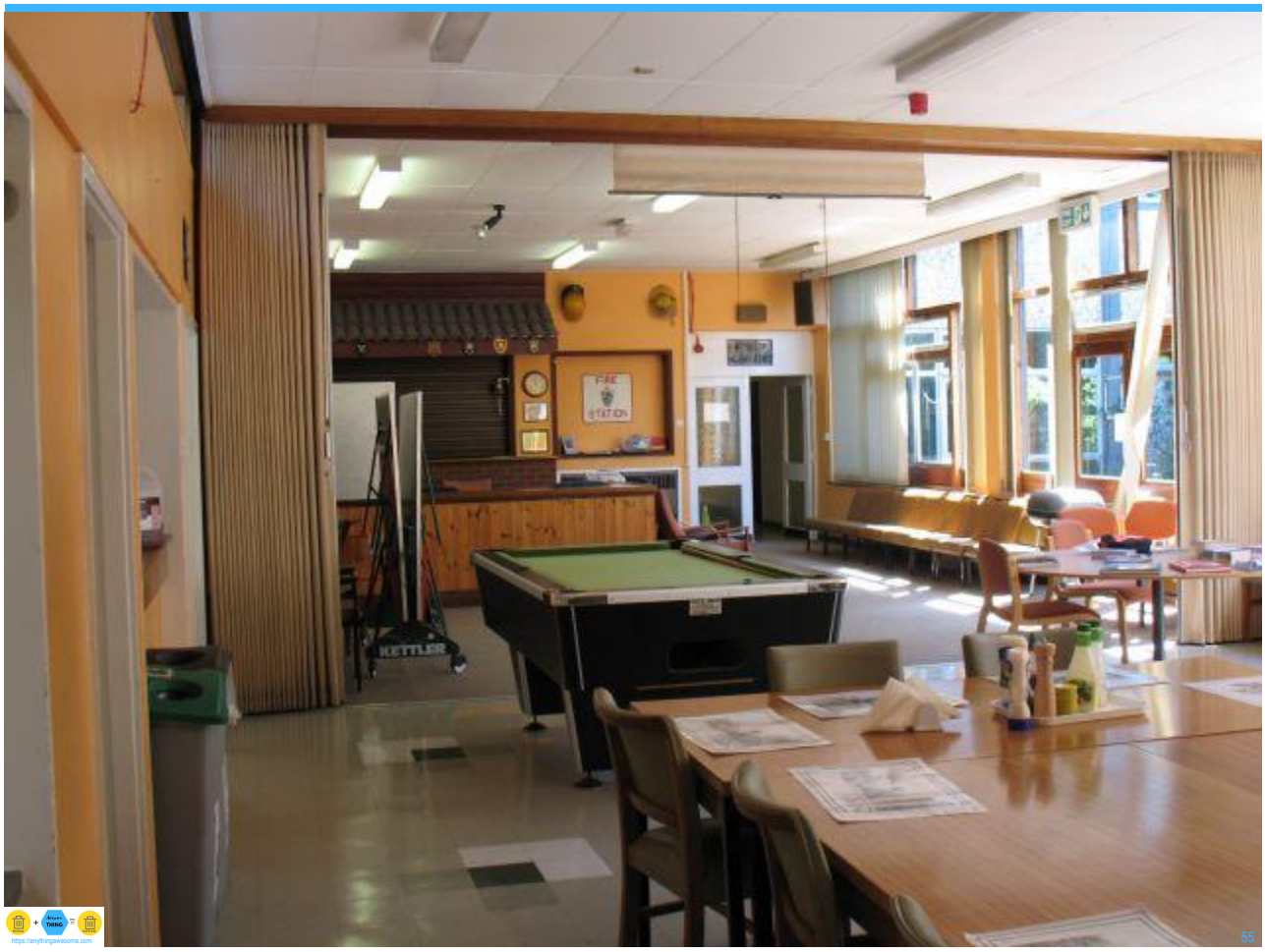
7. Data Governance Policies and Standards
- The Data Governance program will develop and maintain policies and standards related to:
- Data quality
- Data security
- Data privacy
- Data access
- Data retention
- Data dictionary and metadata management

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.

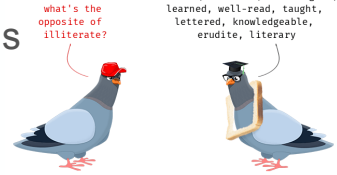
Data and Duct Tape

MacGyver



Full versus Part time

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



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.

Current approaches are not and have not been working

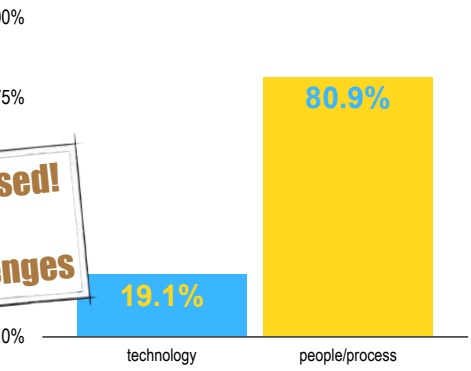
What We Learned From Top Execs About Their Big Data And AI Initiatives

Randy Bean Contributor
CIO Network Contributor Group
Enterprise & Cloud

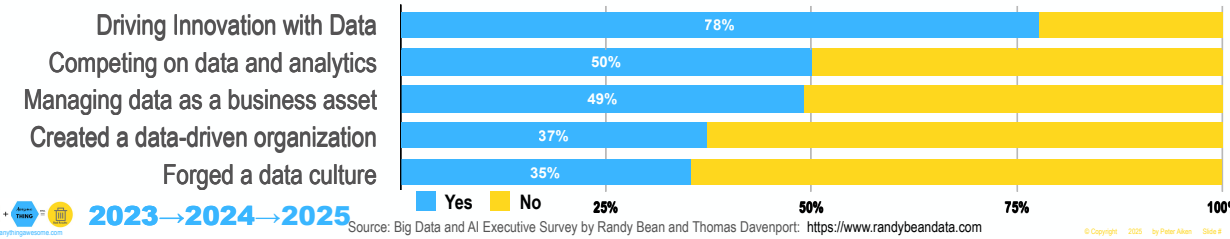
Culture still eats strategy for breakfast

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



2018

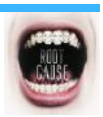


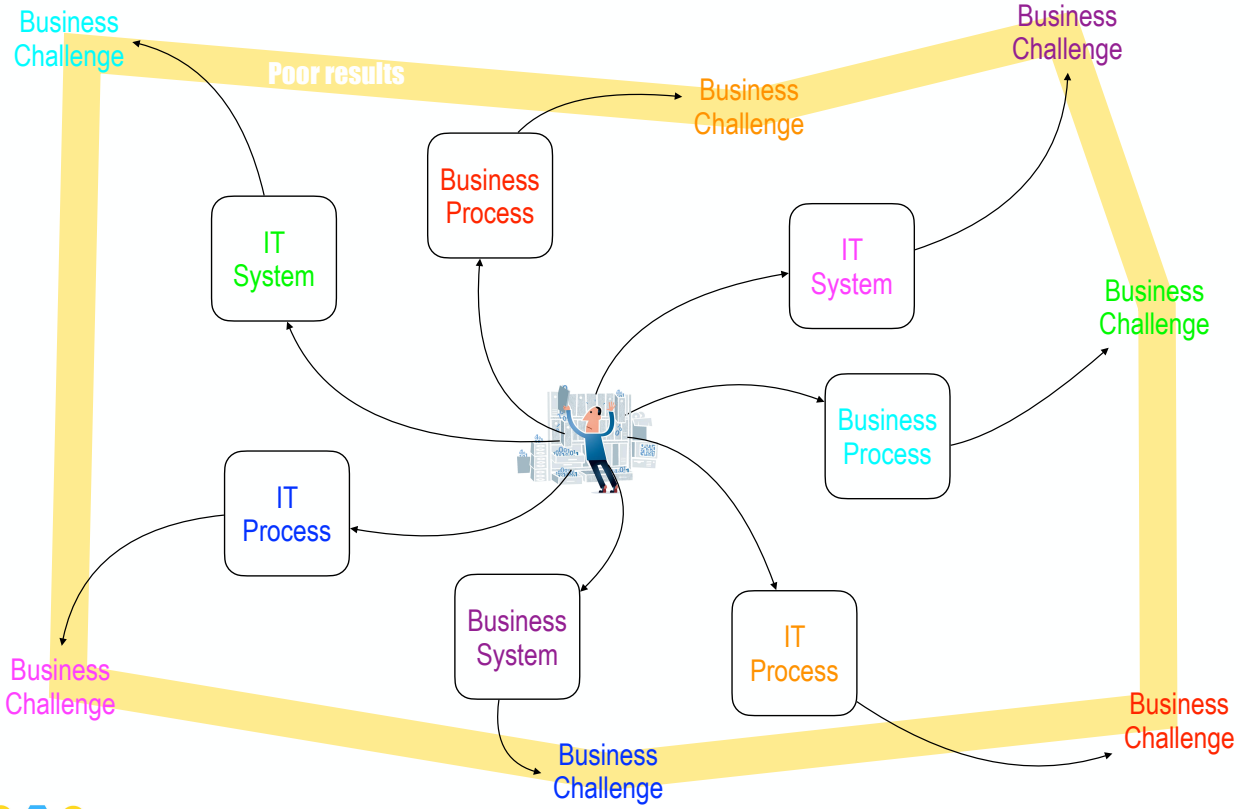
2023→2024→2025

Source: Big Data and AI Executive Survey by Randy Bean and Thomas Davenport: https://www.randybeandata.com

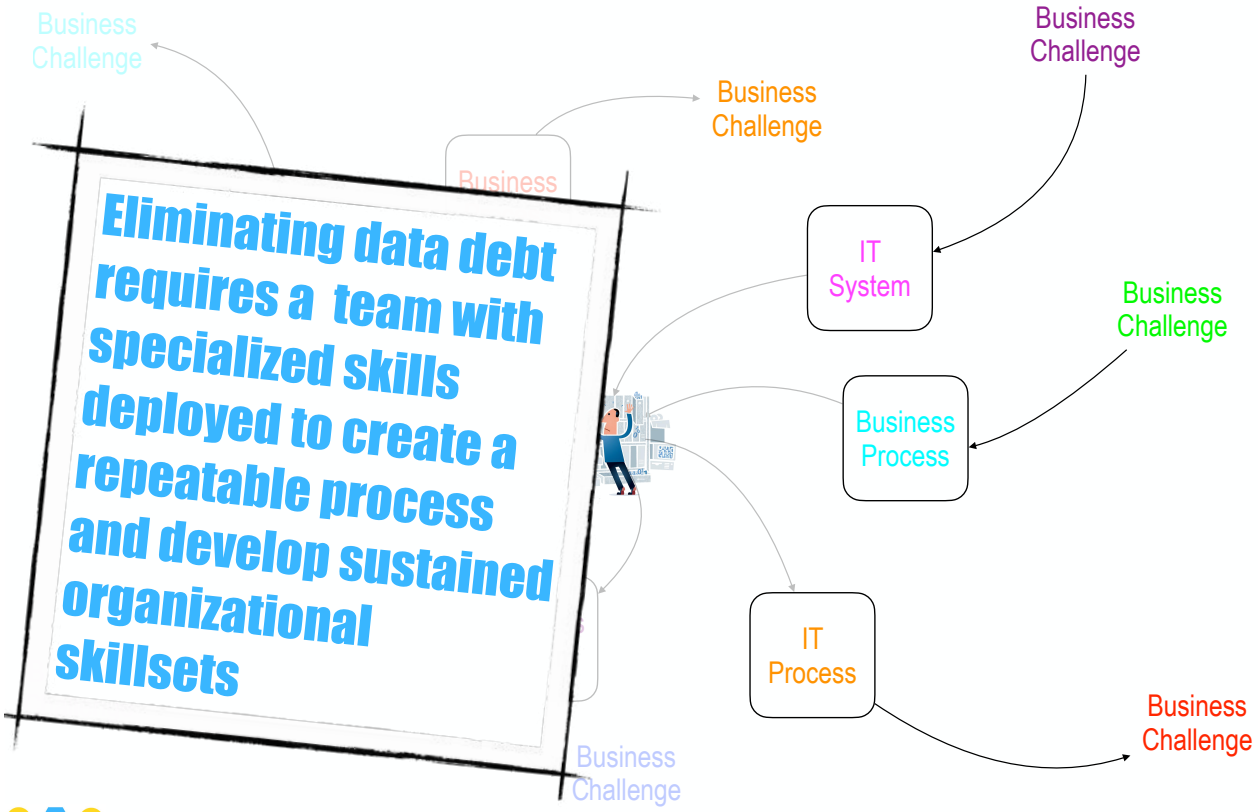
Poor data manifests as multifaceted organizational challenges

Root cause analysis is part of data stewardship

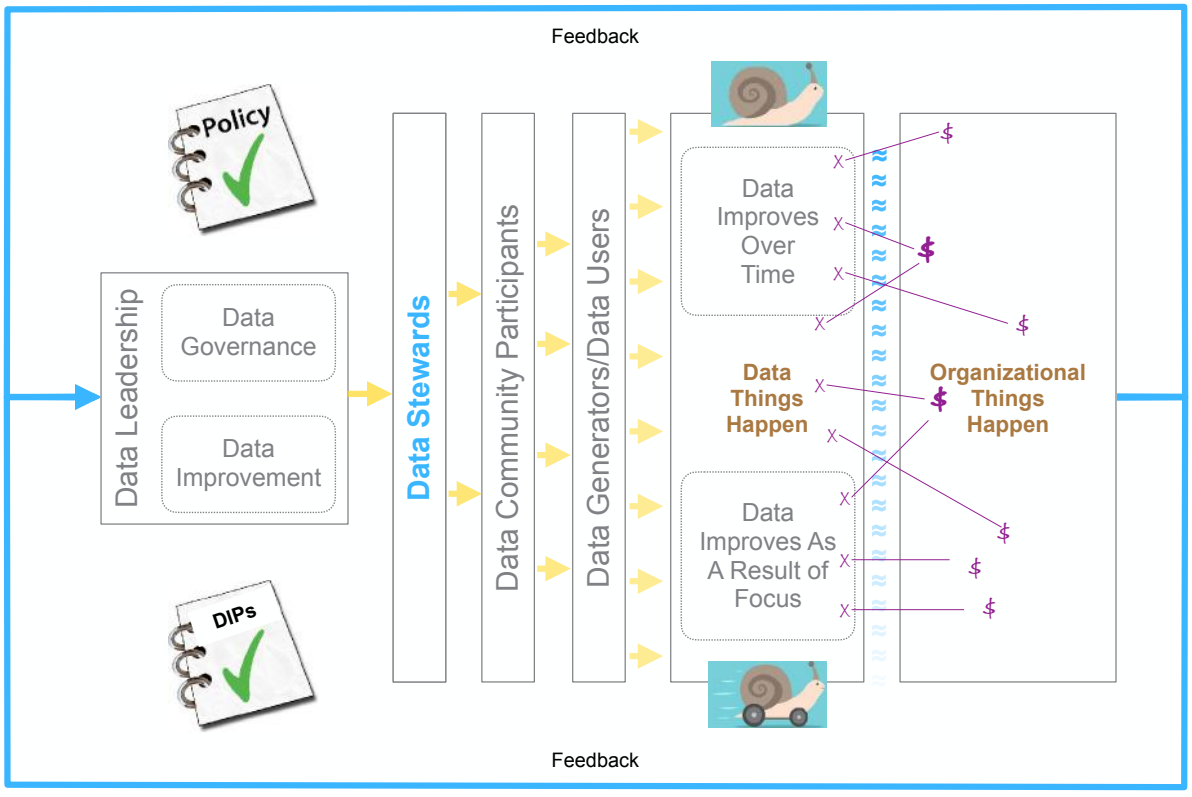




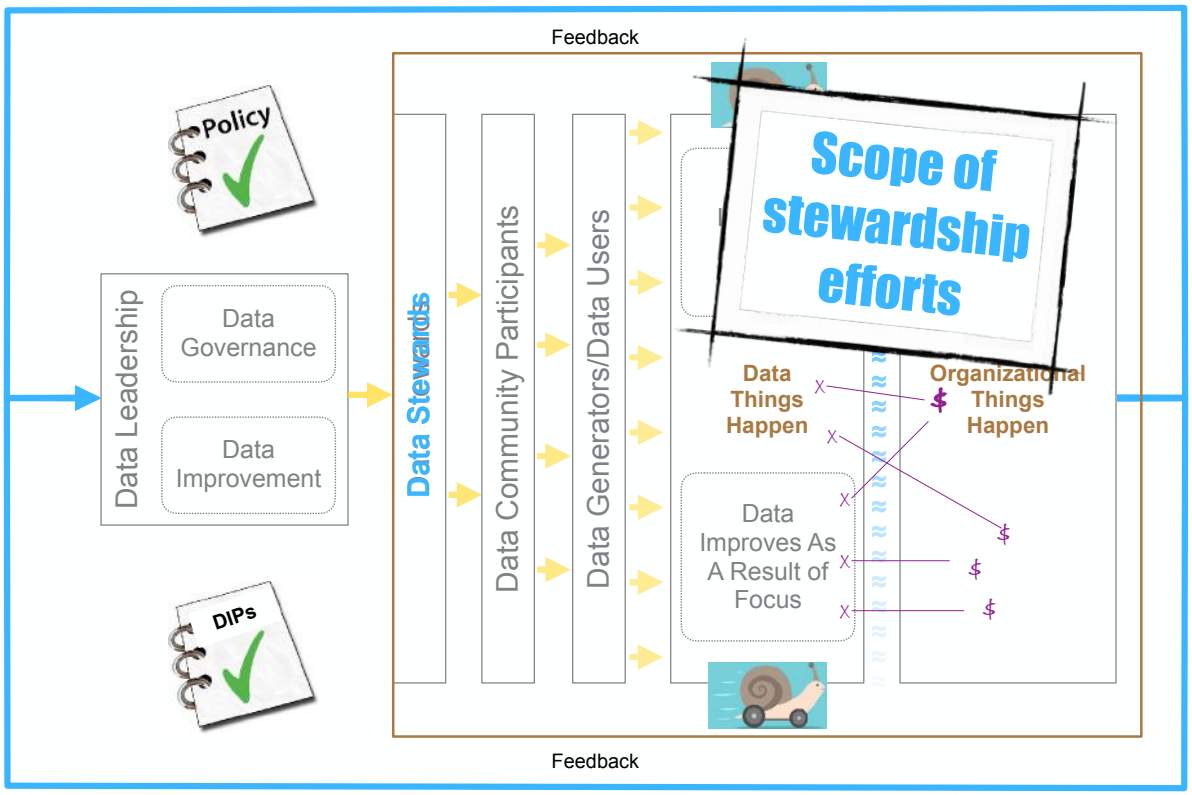
Consistency Encourages Quality Analysis



Data Governance Role: **Produce systemic organizational changes that impact data and work practices over time**



Data Governance Role: **Produce systemic organizational changes that impact data and work practices over time**

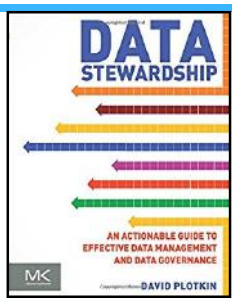


Program Overview

- Why do we need data stewardship as a role?
  - Definitions: Stewardship, data stewards, data debt
  - The role of strategy
  - Data architecture = strategic focus implemented by stewards
- What are they supposed to do?
  - Resolve prerequisite challenges stemming from data debt
  - Data stewardship framework
  - Fire station model (Reactive & proactive foci)
  - Stewardship role (in context of data governance)
- Assigning data stewards (tangible improvement)
  - Start simply
  - Differing cadence (Need for different structural approach)
  - Foundational prerequisites
  - Need for simplicity, agility, practice
- Take aways ➜ Q&A

The Roles of Data Stewards

Data Steward



(list adapted from Plotkin, 2014)

- **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

Data Decisions Variety

Data Source	Steward
• Clinical-domain-specific data definitions	• Departmental clinician
• Master file of IDs (patient, provider, member, etc.)	• IT MDM steward or SME
• Lab codes (LOINC)	• Lab Director
• Pharmacy codes (RxNORM, etc.)	• Pharmacy Director
• Order catalog	• Member of IT staff (usually clinician)
• Diagnostic and procedure codes	• Director of HIM
• Charge master	• Director of Finance

https://www.healthcatalyst.com/why-are-data-stewards-so-important-for-healthcare

Data Steward

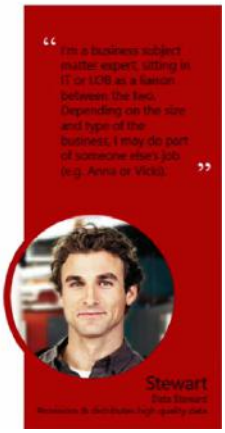
Meet The Data Steward: Self Service Information Management and Data Stewardship

Accountabilities
- Making data useful to the business
- Consistent use of data across the business
- Promoting and achieving high data quality standards
- Resolving data integrity issues across stakeholders

Skills
- 5+ years of industry experience
- Proficient with Office (Excel, Word, PowerPoint). Can learn to use Power Pivot
- Understands data relationships, data process flows. May know SQL.

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

Work Activities
- Analyzes data for quality (particularly as part of BI work), reconciles data issues
- Identifies and acquires new data sources
- Actively analyzes data for 'semantic' quality
- Drives resolution of data integrity issues across business and technical stakeholders. 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

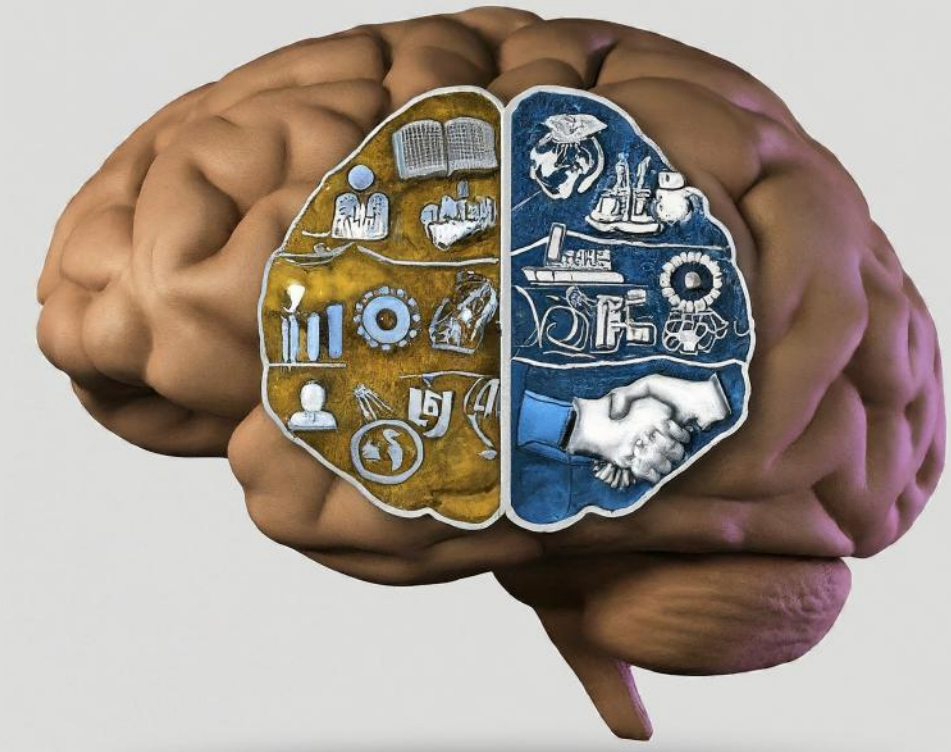


Source: Unknown

What do data stewards do in our organization?

- Tangibly (within their scope)
  - Improve our data asset's value
  - Improve data's use achieving organizational objectives
  - ~~Advocate/evangelize for increasing the scope/rigor of data-centric practices~~
  - ~~Ensure efficient/effective data management practices~~

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)

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.

Responsibility Assignment Matrix

TASK/ ACTIVITIES	PROJECT MANAGER	TEAM LEAD	DEVELOPER	TESTER	STAKEHOLDER
TASK 1	A	R	C	I	A
TASK 2	A	R	C	R	A
TASK 3	C	R	C	A	A
TASK 4	A	A	I	I	C
TASK 5	A	R	I	R	C

Responsible (R)

- This element designates the individual or department responsible for executing tasks. Each task must have at least one person assigned to ensure it is completed and to avoid delays or incomplete work. When it comes to big tasks with significant resources required, multiple people or teams may be needed to ensure smooth execution.

Accountable (A)

- This group holds the authority to make decisions and oversee the tasks within the project. Usually, this person is superior to the responsible team and has a major impact on the project's success or failure. There should only be one person accountable for approvals to avoid confusion and ensure clear decision-making.

Consulted (C)

- Consulted individuals or groups are those whose opinions are sought to help ensure a smooth execution. Those responsible for carrying out assigned tasks should seek advice from this team to handle issues quickly and effectively.

Informed (I)

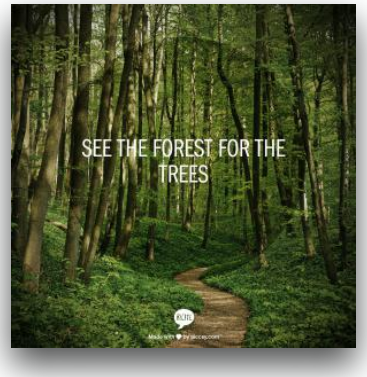
- This section involves keeping relevant individuals or departments informed about various aspects of the project, such as progress, costs, resources, and quality. The person (or people) responsible for execution must communicate these updates to the necessary stakeholders.

Understanding Systems Thinking

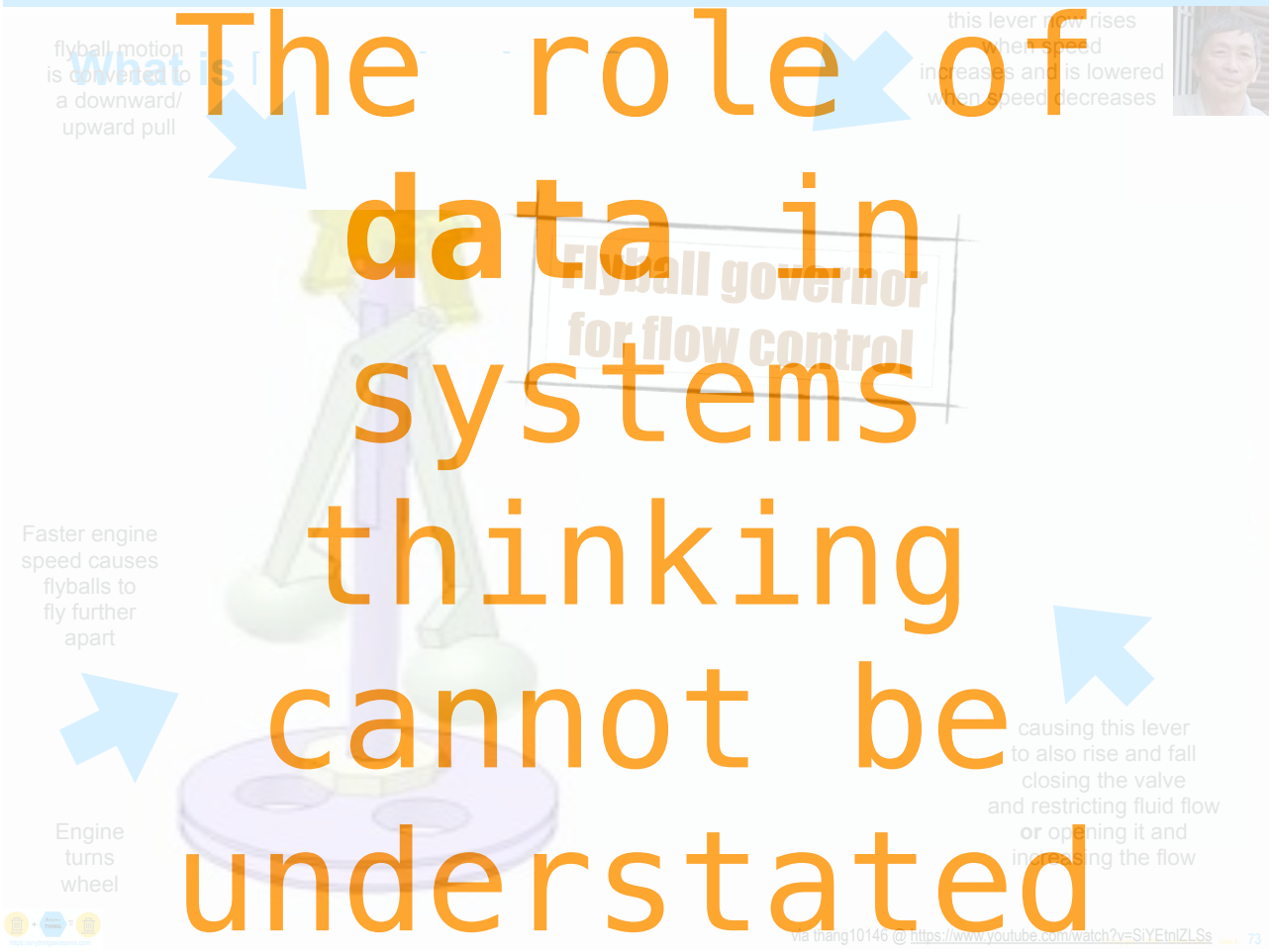
- *A framework that is based on the belief that the component parts of a system can best be understood in the context of relationships with other systems, rather than in isolation.*

- *The only way to fully understand why a problem or element occurs and persists is to understand the part in relation to the whole.*

Capra, F. (1996) *The web of life: a new scientific understanding of living systems* (1st Anchor Books ed). New York: Anchor Books. p. 30



The role of **data** in systems thinking cannot be understated



How to determine what to manage formally?

Too much requires expensive and slow bureaucracy ←

→ Too little misses opportunities

All inputs are data

The Organizational Data Machine

All outputs are data

Interoperability is the primary value determinant

A Data Steward's Goal/Process

Too much requires expensive and slow bureaucracy ←

→ Too little misses opportunities

All inputs are data

Known

Unknown

All outputs are data

Interoperability is the primary value determinant

A Data Steward's Goal/Process

Too much requires expensive and slow bureaucracy ←

→ Too little misses opportunities

All inputs are data

Known

Unknown

All outputs are data

Interoperability is the primary value determinant

A Data Steward's Goal/Process

Too much requires expensive and slow bureaucracy ←

→ Too little misses opportunities

All inputs are data

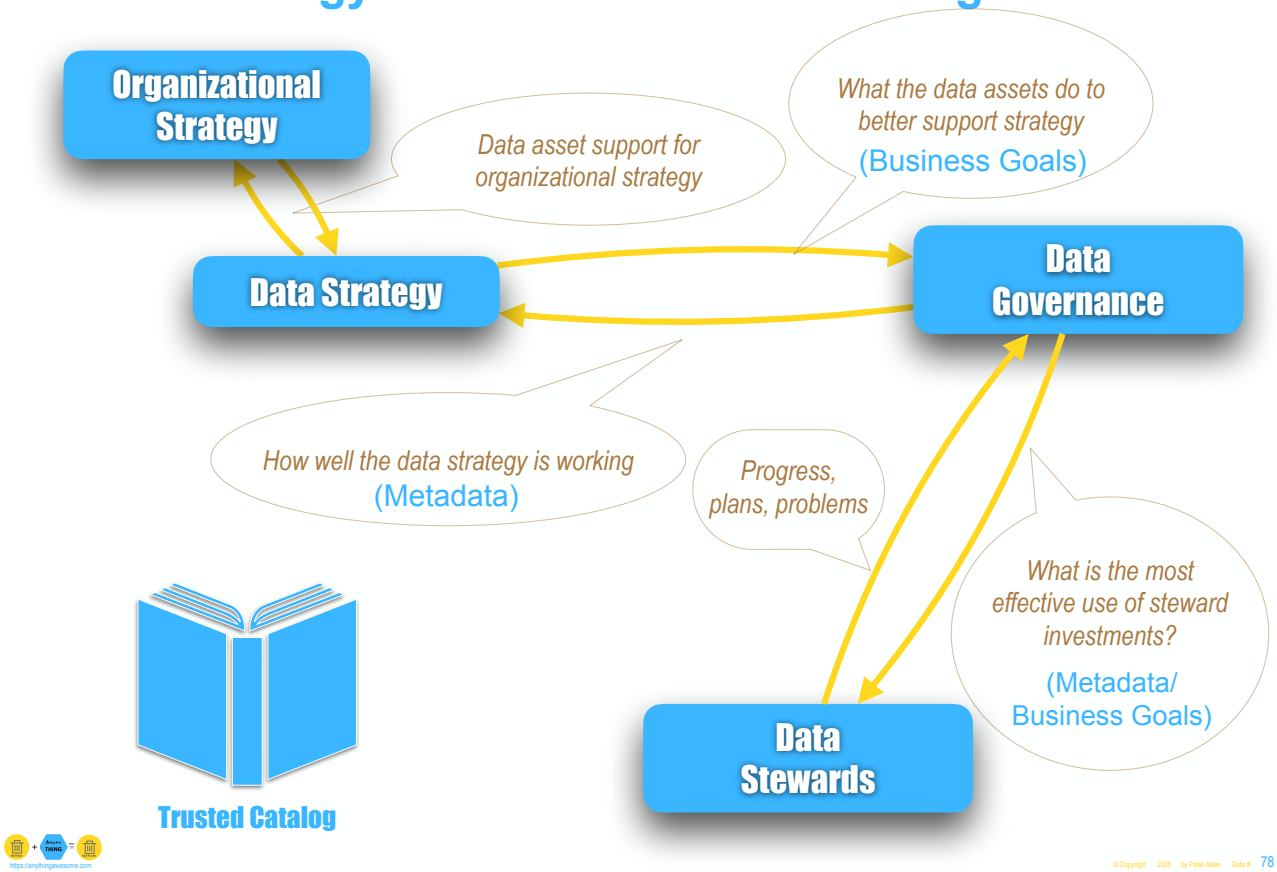
All outputs are data

Known

Unknown

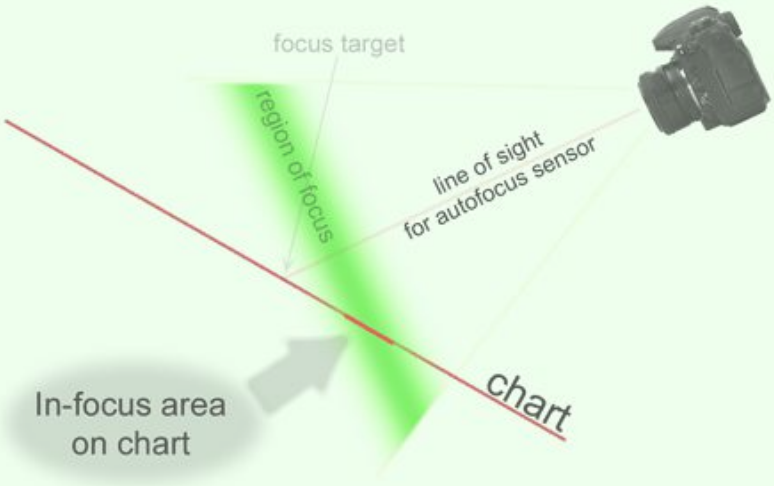
Interoperability is the primary value determinant

Data Strategy and Governance in Strategic Context

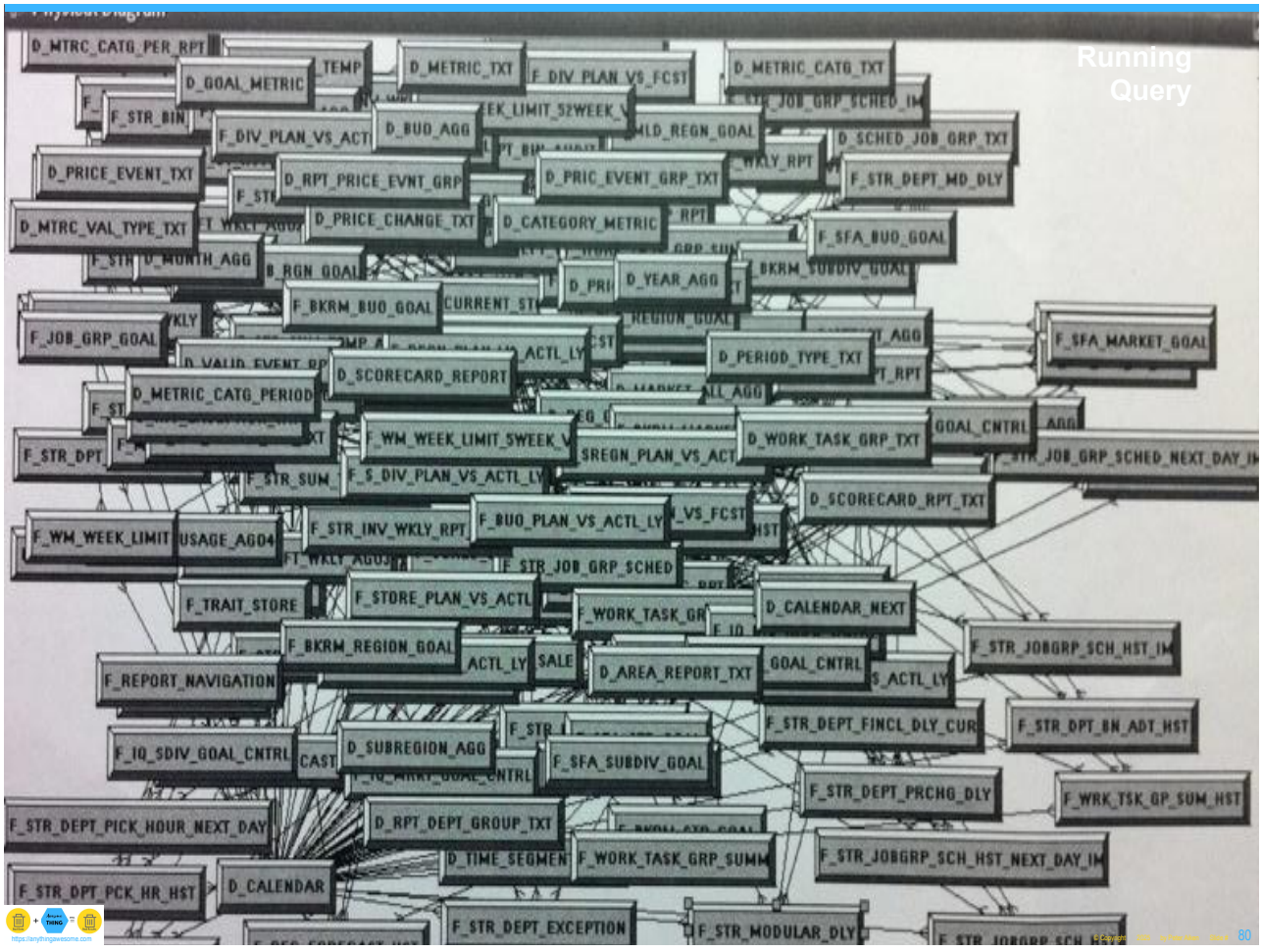


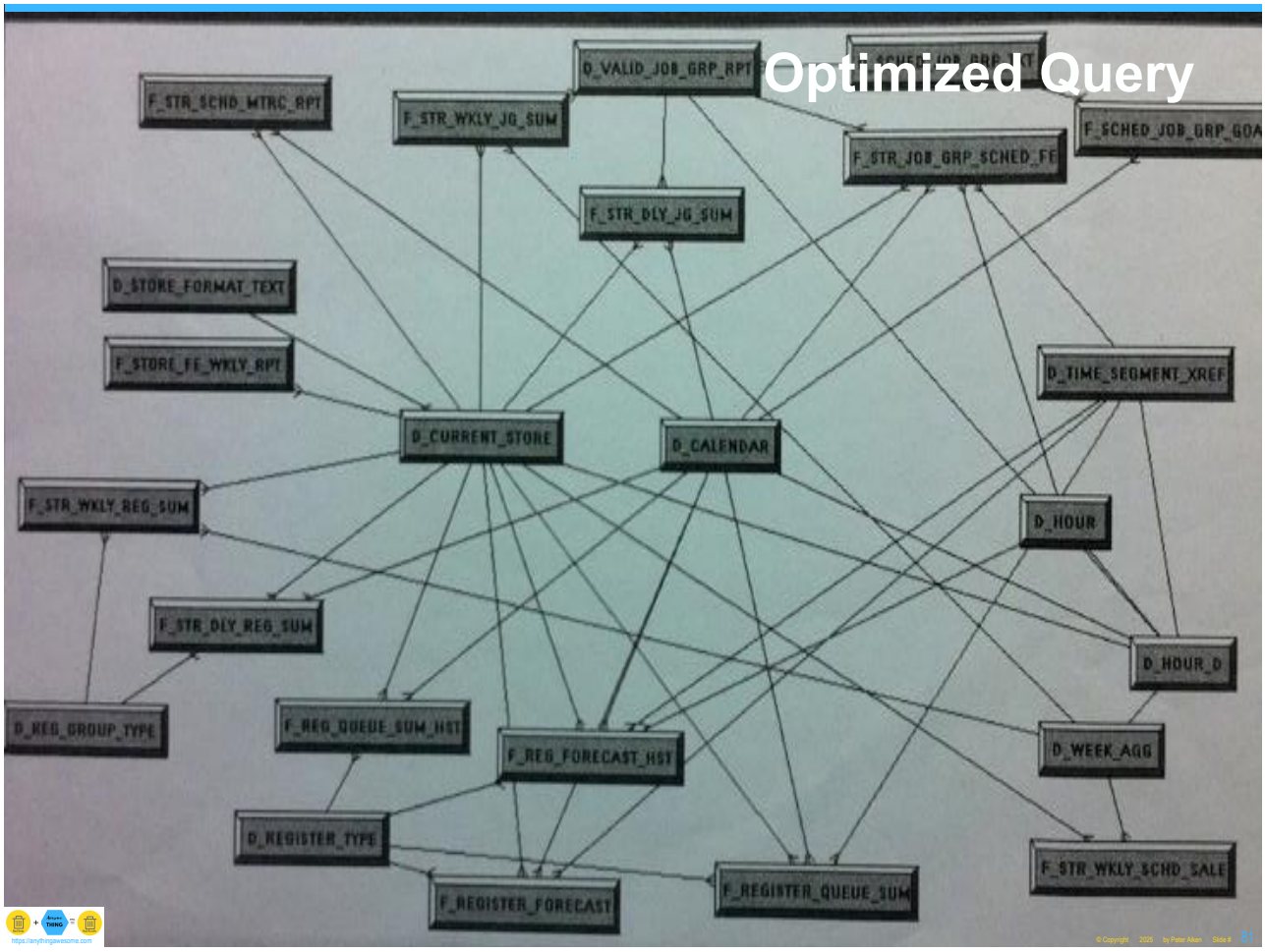
Keep the proper focus

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



- Either way document why!





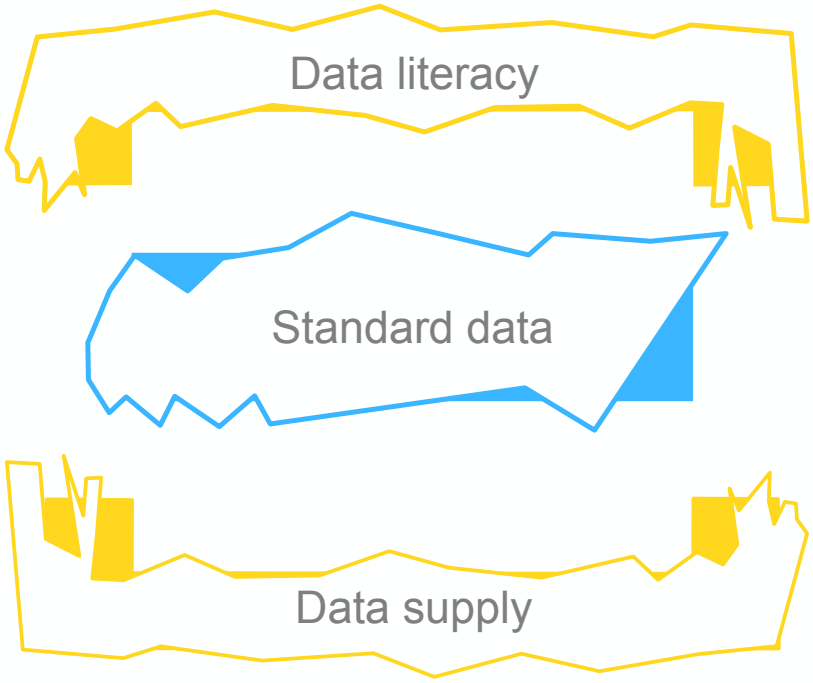
Data Footprints

- 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 ...

bleeding unnecessarily from a lots of cuts

Leverage point - high performance automation



Leverage point - high performance automation

Data literacy

Standard data

Data supply

Leverage point - high performance automation

This cannot happen without investments in engineering and architecture!

Data literacy

Standard data

Data supply

Quality engineering/ architecture work products do not happen accidentally!

Leverage point - high performance automation

This cannot happen without investments in data engineering and architecture!

Data literacy

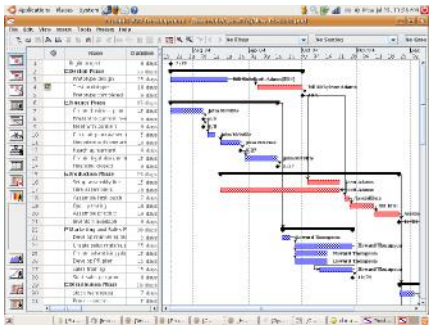
Standard data

Data supply

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

Data is not a Project

- Durable asset
  - An asset that has a usable life more than one year
- Reasonable project deliverables
  - 90 day increments
  - Data evolution is measured in years
- Data
  - Evolves - it is not created
  - Significantly more stable
- Readymade data architectural components
  - Prerequisite to agile development
- Only alternative is to create additional data siloes!



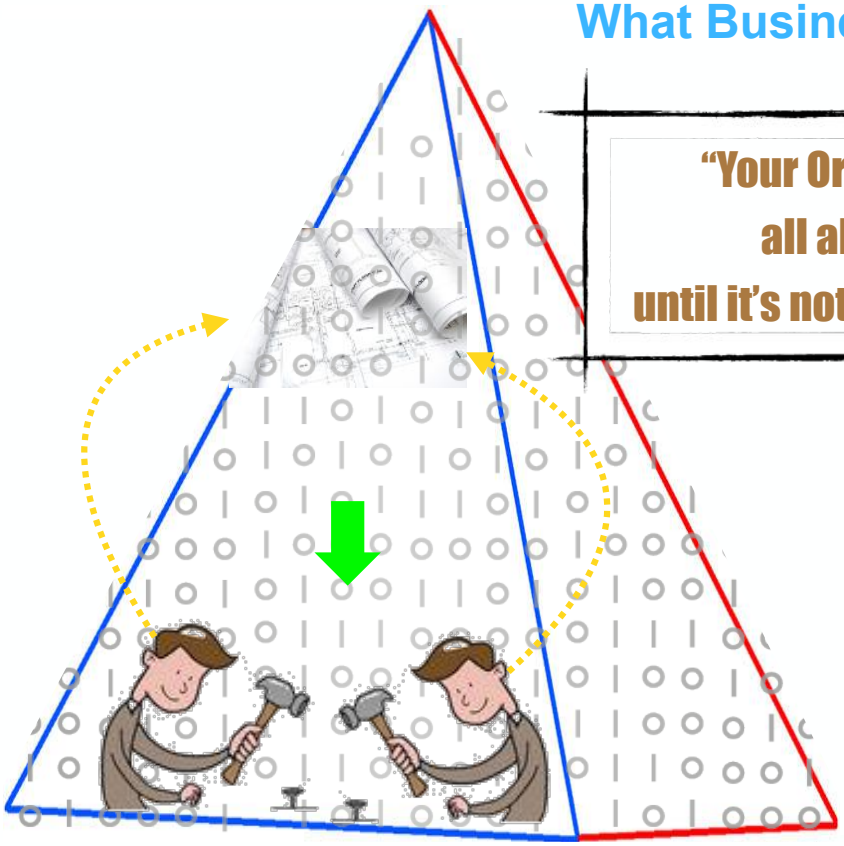
Differences between Programs and Projects

- Programs are Ongoing, Projects End
  - Managing a program involves long term strategic planning and continuous process improvement is not required of a project
- Programs are Tied to the Financial Calendar
  - Program managers are often responsible for delivering results tied to the organization's financial calendar
- Program Management is Governance Intensive
  - Programs are governed by a senior board that provides direction, oversight and control while projects tend to be less governance-intensive
- Programs Have Greater Scope of Financial Management
  - Projects typically have a straight-forward budget and project financial management is focused on spending to budget while program planning, management and control is significantly more complex
- Program Change Management is an Executive Leadership Capability
  - Projects employ a formal change management process while at the program level, change management requires executive leadership skills and program changes driven more by an organization's strategy and is subject to market conditions and changing business goals

Your data steward program must last at least as long as your HR program!

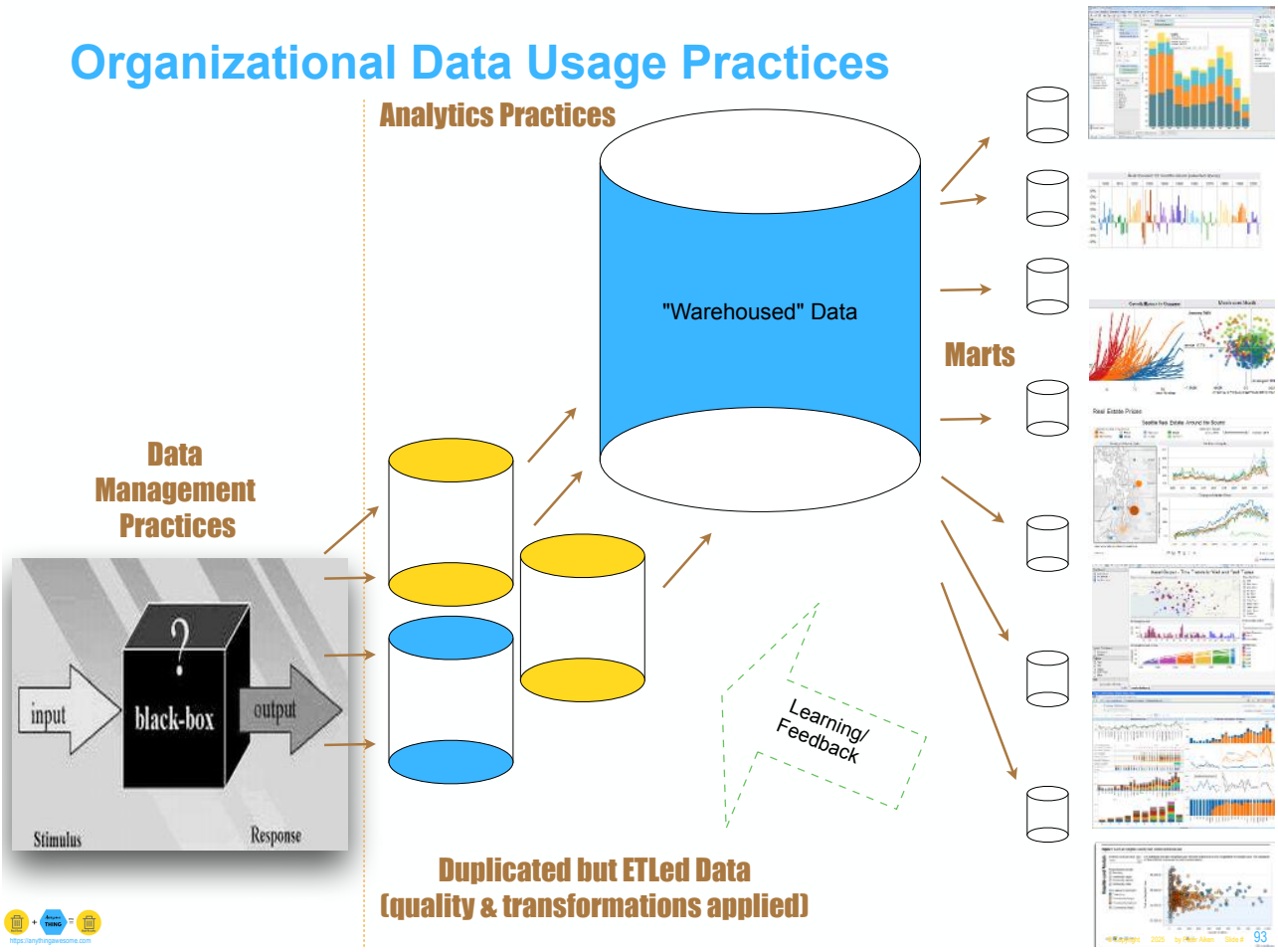
Adapted from http://top.idownloadnew.com/program_vs_project/ and http://management.simplicable.com/management/new/program-management-vs-project-management

What Business are you in?



"Your Organization is all about Data, until it's not about just Data"

Organizational Data Usage Practices



The focus of data stewards should be sequenced

notebooklm.google.com

Welcome to VCU's Policy Program

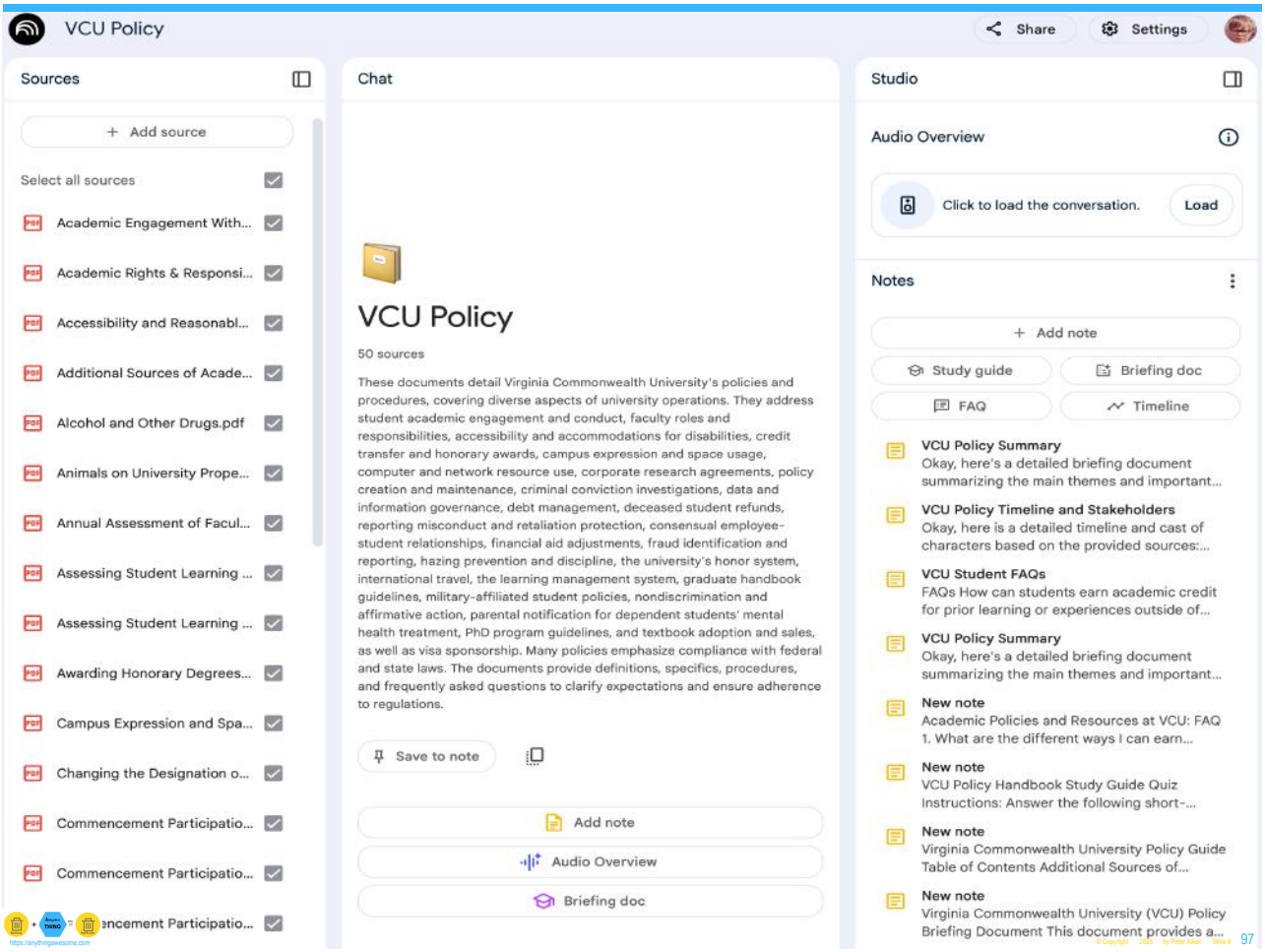
This page houses both the policy library which contains the most current version of VCU policies and resources for developing and revising policies at VCU. This resource is provided as a service to members of the VCU community and the general public. If you encounter any issues accessing pages on this 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.

Search for documents

Found 162 record(s)

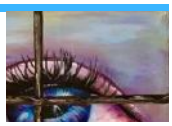


Program Overview

- Why do we need data stewardship as a role?
  - Definitions: Stewardship, data stewards, data debt
  - The role of strategy
  - Data architecture = strategic focus implemented by stewards
- What are they supposed to do?
  - Resolve prerequisite challenges stemming from data debt
  - Data stewardship framework
  - Fire station model (Reactive & proactive foci)
  - Stewardship role (in context of data governance)
- Assigning data stewards (tangible improvement)
  - Start simply
  - Differing cadence (Need for different structural approach)
  - Foundational prerequisites
  - Need for simplicity, agility, practice
- Take aways ➜ Q&A

Data Stewards: Defining and Assigning

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



External Comprehension

Everything Else Data | **Data Management** | **Data Governance Program**

Data (blah blah blah)

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

Data Program

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)

Take Aways

- Since stewards own the results
  - They control the remediation process
- Need for professional data stewards is
  - Increasing
  - Data volume is increasing
  - Lack of practice improvement
- Data stewardship is a relatively new discipline
  - Must conform to organizational constraints
  - No one best way
- Data stewards
  - Must be driven by a data strategy complimenting organizational strategy
  - Direct data management application
  - Speak the language of metadata
  - Should drive process improvements

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

Upcoming Events

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

Reference Data vs Master Data Management

8 April 2025

The Importance of Ethics in Data Governance

13 May 2025

Implementing Effective Data Governance: A Practical Guide

10 June 2025

Brought to you by:

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

DATAVERSITY

Critical Design Review?

Mentoring?

Collaboration?

Executive Data Literacy Training?

Independent Verification & Validation

Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

Use your data more strategically?

Tool/automation evaluation?

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

Chapter Contents

Chapter Overview

This database ain't big enough for the two of us

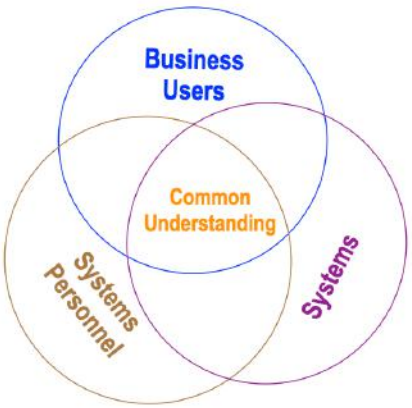
– Bumpersticker seen on an automobile in Texas

The bumpersticker should really have stated "There is no database big enough for two bosses." Importantly, 1) this has always been true, and 2) it means absolutely nothing to most of the public or much of IT. Let's address each of these separately.

Just as in any situation where coordination, integration, and information are required, there must be one and only one individual implementing decisions to maintain integrity, continuity, and operational capabilities. Required minimally from a change management perspective, this can always be used to justify DG in general. Ask the skeptical: how can any complex adaptive system function with multiple Chiefs?

The public and unfortunately too many in business and IT do not understand this sort of basic law of (data) nature. Because they are not data literate, when someone proposes having multiple chiefs for database operation, or that group X should 'own' dataset Y, or that the DG group should report to the CIO–they do not know these are not a workable concepts!

DG is not focused primarily on databases, clouds, or other technological ephemera. Instead the DG framework must be understood identically by business users, systems personnel, and the systems themselves (as shown to the right). This essential, metadata-based communication is at the heart of any enterprise operation. DG removes barriers to data efficiencies, allowing organizations to function more effectively and efficiently. Resources consumed by bad data practices can now be used to support the mission.



Increasingly organizations are attempting to do 'more' with data. This represents the other strategic dimension, innovation. By definition, most attempts to innovate will fail–so the lessons learned by becoming more effective and efficient, will also help in this innovation dimension. Innovating with data requires programmatic support for the efforts–well supported by data infrastructure and mature organizational data practices.

It is the responsibility of DG programs to manage this and other delicate balancing acts required to successfully contribute to better organizational use of data. DG is a comparatively new, certainly unstandardized, and under-studied topic. While some excellent DG programs are maturing, the majority have not. This leaves individuals and organizations the sequential tasks of:

1. learning about data (and then)
2. learning about their data (next)
3. developing plans to increase the data literacy of their executive leadership (then)
4. their knowledge worker population (before expecting to)
5. make progress faster and further with data.

This chapter takes you through the who, what, where, when, why, and how of DG. It provides a common-basis for building individual and organizational knowledge of this topic–starting with the **why**–the motivation for DG–followed by the **who**, **when**, and **where**. The **how** section is a bit longer and the bulk of the remaining material concentrates on the **what**–a way to successfully start to govern subsets of your data.

Most organizations should not attempt to govern all of their data. Successful DG program goals include subsetting their data into essential and non-essential data. Governing the essential subset and ignoring (or better still removing) the rest reduces the size of the challenge. Since the definition of an organization's essential data will differ from organization to organization, the governed data will also differ among organizations.

One quick word about the use of the term **bespoke** in the title. It is of course deliberate. The only way that **your** organization can use data to better support organizational strategy, is to use **your** data in support of **your** strategy using the capabilities that **you** currently have. Cookie cutter methods will not help **your** organization learn about **your** data!

Why does data need to be governed?

A friend was speaking with an organization on data matters and noticed that the urinals in the restrooms all had unique numbers. Presumably this was in case of malfunction so that the specific instance could be more rapidly identified. Of course my friend used a suitable-for-work (as opposed to not-suitable-for-work) photograph to make a point to leadership that (at least for this organization) it was worthwhile to keep

maintenance histories of this equipment type. Ironically, it was noted that the substance of the discussion for which my friend had been invited was whether the organization should maintain similar information about their organizational data assets. The photo provoked a nice motivational discussion with a decision to proceed with DG as the outcome. After all, if we are going to govern our restroom facilities, shouldn't we also govern our data assets?

Writing as a deeply, industry-immersed university professor, I can say that the academic community has failed its customers with respect to integrated data knowledge. For generations we have graduated students who have become leaders in business and IT. The only class taught about data was really about database development. Smart students who placed their trust in the educational system, were educated that the only concept they needed to learn about data was *how to build new relational databases!* No one should be surprised that one of the major DG challenges is that far too many poorly designed databases clutter most organizations or (more increasingly) their clouds. As Abraham Maslow stated: "**If the only tool you know is a hammer, every problem looks like a nail**."

When considering the asset itself, data has a unique collection of properties including the following from Doug Laney. Data:

- Does not obey all of the laws of physics
- Is not really visible
- Is non rivalrous (many can use it at once)
- Costs of providing an additional copy are zero
- Is non depleting
- Does not require replenishment
- Is regenerative
- Has low inventory and transportation/transmission costs
- Is more difficult to control and own that other assets
- Can be eco friendly
- Is impossible to clean-up if you spill it.[1]

When considering career fields and learning experiences, not all data professionals take similar paths. For example, data scientists often **discover** useful data maintenance utilities instead of **learning** that various classes of tools exist and when to apply each as part of their educational programs. For many, data is like the story of the blind men and the elephant and collectively it is DG responsibility to shape this understanding into an organization-wide perspective.

For these and other reasons there continues to be questions as to whether data processing should continue to be part of IT or of the business or o special operations

[1] See **Datanomics** by Doug Laney Routledge Publishing 2017 ISBN 1138090387

such as finance and risk? While the Federal Government resolved this issue correctly with new FEPA legislation, the jury is still out on the rest of the world. Currently it is comprised of ⅓ of each type: one-third reporting to CIOs; ⅓ reporting to CEOs; and ⅓ reporting to CFOs/CROs.

Long lasting consequences of poor data decisions?

Unfortunately, short-term application-centric thinking[2] has dominated, relegating development of data products to subsets of ERPs, digitization initiatives, or cloud hosted projects (to name just a few types). Virtually none of the popular software integration packages from the major vendors have escaped the long-term consequences of inadequate data Design (big 'D' is used to emphasize the entire lifecycle). These well documented imperfections are locked in for life-wrapped as they are, in a dense set of application constructs interwoven with the imperfect data model. Worse still, the corrections to the organization's data and processing are layered on as additional code–complicating the apps still further. The vast majority of database functionality is not used beyond table-handling. In this manner, developers restrict any subsequent data investment benefits and decrease data leverage potentials. At the very least, DG must illustrate and resolve the 20-40% of IT budgets that are devoted to data evolvution:

- 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)

None of these are accounted for in the usual (and very important) **data storage costs–measure**. DG must also articulate these various costs and tradeoffs associated with increased data rigor (or the risks of not doing so) to the rest of the organization.

Mounting data debt

The failure to do any of this has caused organizations to pay to accumulate large amounts of data debt. (Yes, the indignity that your own organization is creating data pollution that is directly harmful to its operation should be professionally embarrassing!) It is not easy to visualize the cost of data debt but the phrase *many many many unnecessary paper cuts* describes the situation well. Data debt slows DG efforts making everything slower, of lower quality, cost more, or present increased risks.

Data debt is like quicksand that mires down all efforts. Defined simply, data debt is: the time and effort it will take to return your data to a governed state from its likely current state of ungoverned. A quick back of envelope calculation of data debit can be

[2] See **The Data-Centric Revolution: Restoring Sanity to Enterprise Information Systems** by Dave McComb Technics Publications ISBN 1634625404

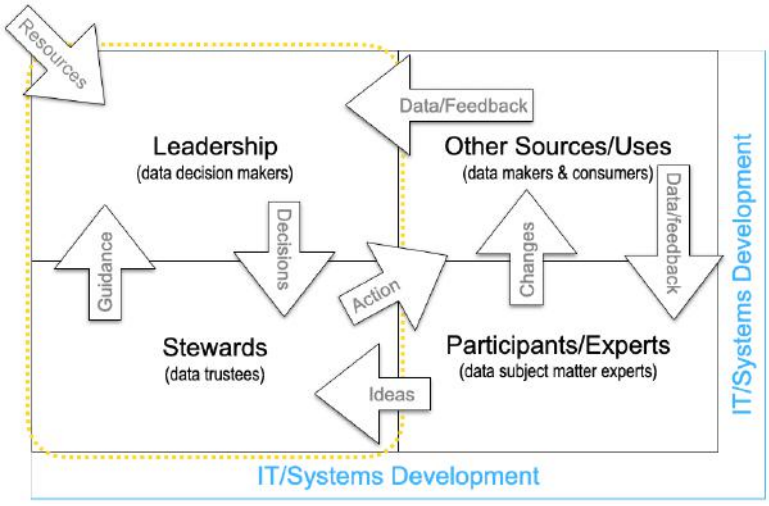
done using the data storage costs that are perhaps the most tangible and objective data measure. At least 20% of that data is redundant, obsolete or trivial (or ROT).

The good news about finding and eliminating data debt is that things can get faster, better, or cheaper. The bad news is that new skillsets are required of the DG team and that diagnostic and analytical systems thinking still requires annual proof of value. The knowledge-base of grey-beards who know how to apply these skills is shrinking as these individuals are judged expensive and encouraged to retire.

In summary, data needs to be governed because society was not taught that it required specific treatment until it was too late. Because individuals do not know that they do not know, it has been difficult to educate them to the need. By focusing on concrete results, organizations have better success making the case that an investment in DG will benefit the organization in specific measurable ways.

Who needs to be involved in DG?

Unfortunately at many organizations, **everyone** has been responsible for data quality and this approach has produced the current unsatisfactory state. It is critical to start DG educational efforts with executives because 1) they are willing to invest in learning and 2) their data decisions have the greatest impact on the organizational data practices. The next goal for all DG programs is to also increase the data literacy of all organizational knowledge workers.



As illustrated, DG efforts are generally built on an IT provided support/foundation/ infrastructure. A leadership component provides resources and clears barriers for the effort. Primary functions are (ideally full-time) data stewards who provide guidance and design/implement decisions. Typically these two groups form the basis for DG organizations. Also, highly involved (and incorporated) are various SME or subject matter experts who know the required data and processing details. Then of course there is everyone else. As noted, DG efforts need to be integrated with both organizational and IT governance.

When is it appropriate for organizations to invest in DG?

By now I hope that you agree this is a silly question. The 20-40% of IT costs (referenced previously) is easily gauged. As the maturity of the DG practice matures,

processes can be optimized for key operations. By keeping disciplined measures, organizations have developed expertise in these practices. Keeping the focus on a integrated full time team permits the case to more easily be made when timing investment in a second or third DG team.

Digital and data are dependent on high speed automation/data processing that requires significant amounts of organizational data literacy, data standards use and quality data supplies. Continue to evaluate and evolve DG frameworks to refine the organizational focus. Over time this approach should evolve into the standard Deming plan, do, check, act (PDCA) cycle.[3] An incomplete list of potentially useful standards that can be created with the required measurable controls is listed below.

- Access standards
- Change management
- Security
- Storage
- Reporting
- Classifications
  - ➡ Secure
  - ➡ PII
  - ➡ Competitive advantaged
  - ➡ Public

Where should organizations get started with DG?

DG is a rare triple benefit capability that helps refine data strategy, improve the quality of the players, and improves data used to support the mission. However, getting started with DG can be and has been accomplished by a moras of ill-defined and vendor specific methodologies–most of which have no reported research results.

An easily understood model (the theory of constraints[4] or TOC) views programmatic data support as a manageable system. The system is limited in achieving more of its goals by a small number of constraints. There is always at least one constraint, and TOC uses a focusing process to identify the greatest constraint and restructure the rest of the organization to address it. TOC adopts the idiom that "a chain is no stronger than its weakest link," and processes, organizations, etc., are vulnerable because the weakest component can damage or break them and adversely affect the outcome.

Key is to visualize the various data flows through the organization and understand the value of controls in relation to various processes, risks, outcomes, and performance. The costs of various blockages can be ranked and estimated. What changes made at the data level could most help the organization achieve its strategic goals?

Iterative problem solving provide additional benefits beyond challenge solutions. Team problem solving enables increased organizational data literacy and some go as

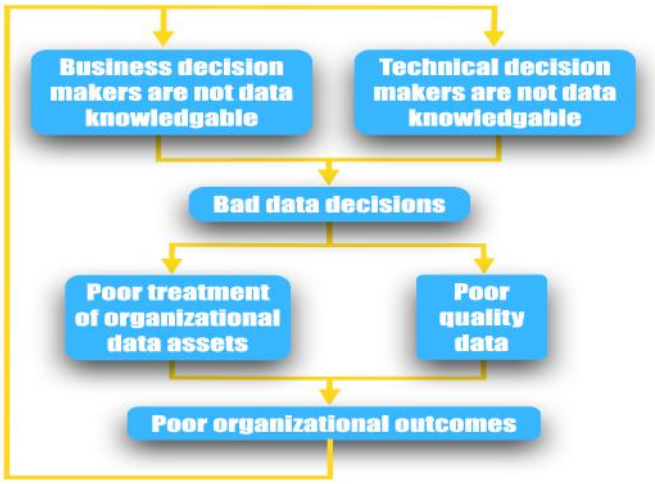
[3] https://en.wikipedia.org/wiki/W._Edwards_Deming#PDCA_myth

[4] https://en.wikipedia.org/wiki/Theory_of_constraints

far as considering these capabilities their 'secret sauce.' It just makes sense to support a group of individuals who possess knowledge of your data and its uses.

Focus first on organizational strategy. Understanding intricately, the data flow supporting increasing performance, decreasing costs, impacting times, and better managing risks. Identify the various types of organizational challenges sharing the same data or (better still) data errors. These become the focus of the first iteration of a data strategy cycle. It is overseen by the DG program and coordinated to be most collectively helpful to organizational as well as IT strategy. Ensure you complete a full cycle to include feedback/improvement/lessons learned/organizational memory/ change cycle components. Heavily incorporate the use of 'branded' data checklists and standard control development.

And then (as it says on the shower bottle) lather, rinse and repeat. This is really to only way to escape the bad data cycle. IT and business decision makers are not knowledgeable about data and good data practices. They make poor decisions about data that result in poor treatment of organizational data assets and poor quality data. Both of these lead to poor organizational outcomes.



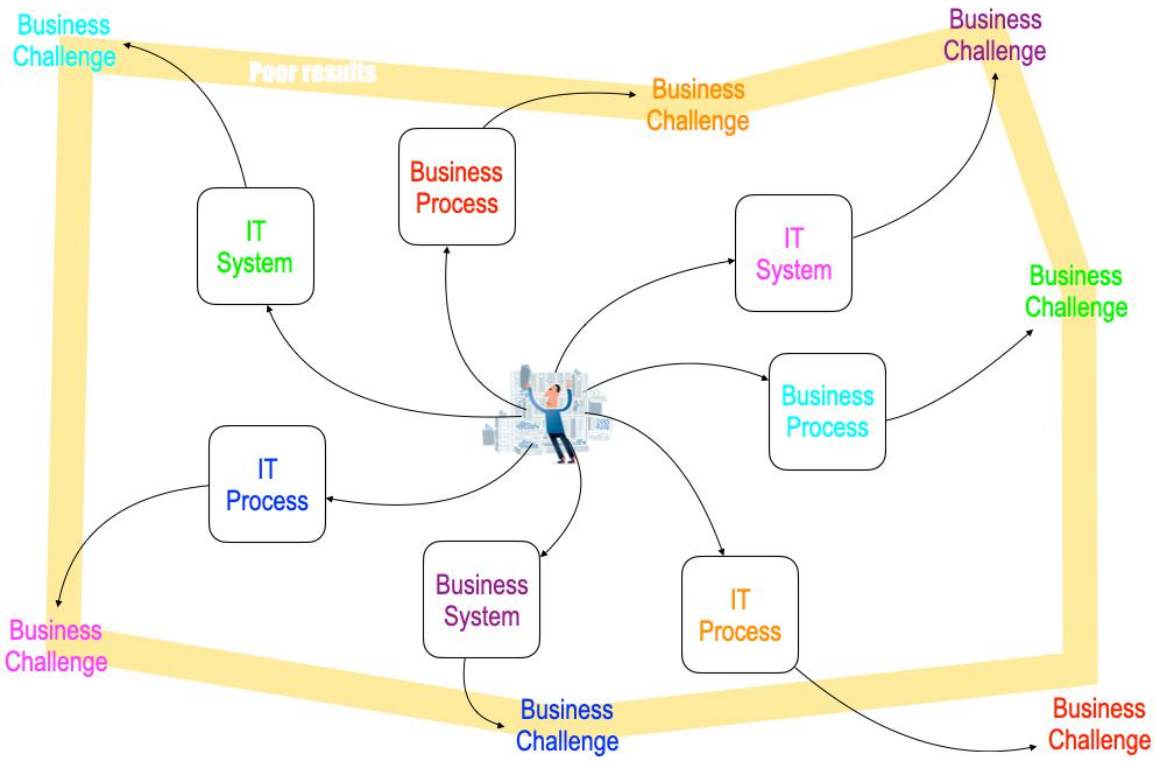
How should organizations apportion their DG efforts over time?

Data debt's impact

Over time, organizational data debt clogs value-adding pathways in a manner similar to the 40% of the internet that is now clogged with malware. Data debt is responsible for inflicting uncounted tiny hidden data factories[5] on organizational performance-making everything cost more, take longer, deliver less, and at increased risk. Eliminating data debt requires a team with specialized skills deployed to create a repeatable process and develop sustained organizational skillsets.

A major motivation for increasing the data literacy of all knowledge workers comes from the fact that most organizational challenges come filtered through various IT and business practice combinations. The reason for multitude of paper cuts, is that the DG challenges are filtered through various business processes and IT systems. As a result, common challenges go unrecognized with each instance requiring treatment instead of correcting the underlying data challenge.

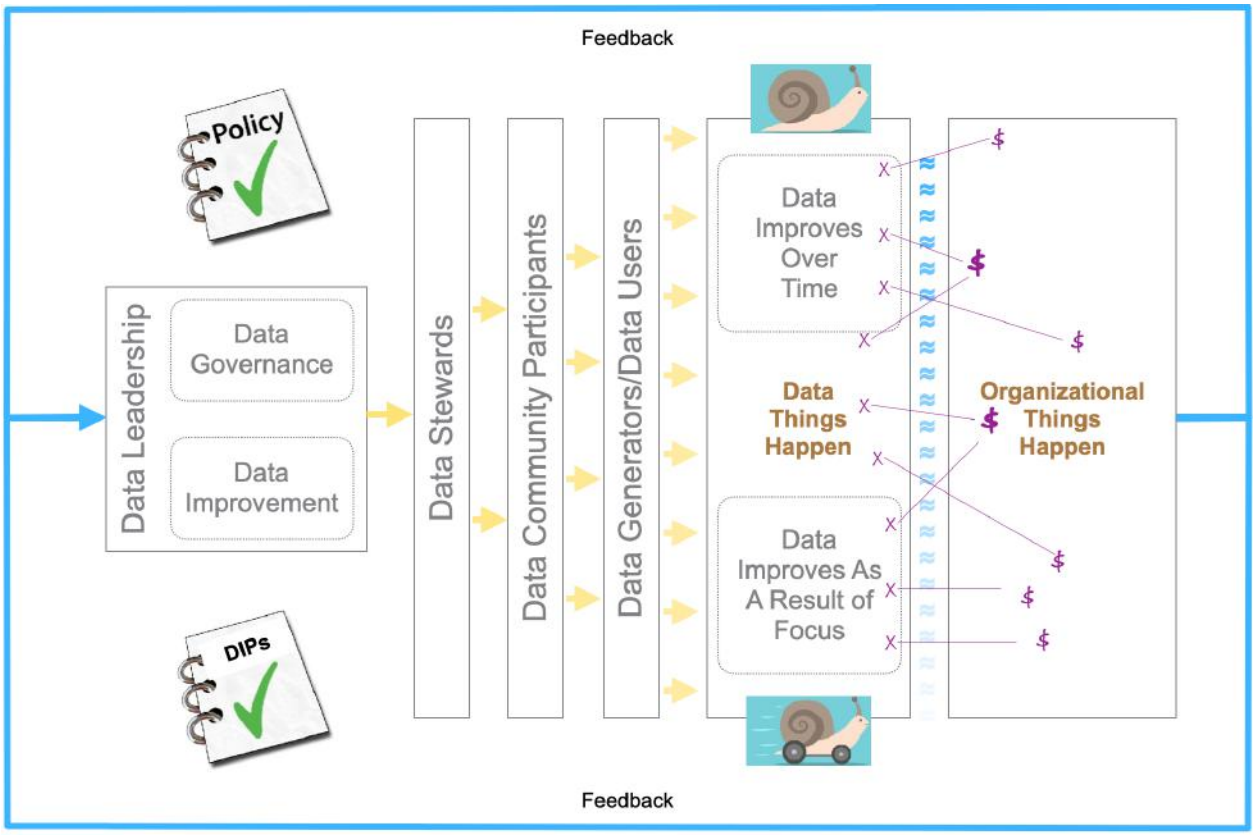
[5] https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year



A key aspect is to evaluate your architectural abilities to build/evolve towards organizational data capabilities in a 3-step process. First, you need to improve the quality of existing organizational data. Too many organizations do not have enough information about the quality of their existing data. These data quality challenges fall into two categories: practice-related data quality challenges and structure-related data quality challenges. Second, the framework must support your efforts to increase the data literacy of literally your entire executive team and knowledge worker population and especially those who already practice data. Finally, only when you have improved your data and your organization's ability to work with data, can you hope to improve the way that data supports your organizational strategy.

Proactive versus reactive DG

One rather traditional realization (almost a rite of passage) is that what ever changes are made to the organizational data practices might take literally years to be able to exploit it. In CIO terms, it can often be a successors, successors, successor that will benefit from DG initiatives. As this realization sets in (that time equals years), DG initiatives come under pressure to 'do something more quickly.' As illustrated, a secondary capability is established to more effectively produce results as a result of direct intervention or Data Improvement Projects (DIPs).



MacGyver abilities

While perhaps not widely acclaimed, the 1980's TV series MacGyver became shorthand for an non-traditional and innovative problem solver who always carried a Swiss Army knife.[6] In the same manner, the DG program must imagine itself as the 'help desk' for organizational data. Literally all data challenge solutions should be minimally coordinated and, in many instances, led by DG. The key is to develop new data capabilities within a dedicated group focused on organizational data governance. Have this group focus on and conquer a series of DG challenges, producing positive ROI numbers.

What organizational needs does DG fill?

It is useful to describe the organizational needs that DG fills. These include:

- Improving the way that data is treated as an asset
- Available but not widely known research results
- Using data to better to support the organizational mission
- Using data strategically

[6] https://en.wikipedia.org/wiki/MacGyver

Improving the ways that data is treated as an asset?

One of the primary challenges for organizations is to learn how data requires specific considerations. If you consider data as an asset (and currently most business leaders do not yet do so) then one should expect that it would be treated as other organizational assets. I use a series of questions developed by my colleague Dr. Christopher Bradley to help organizations determine whether their data is maintained as an asset. They are:

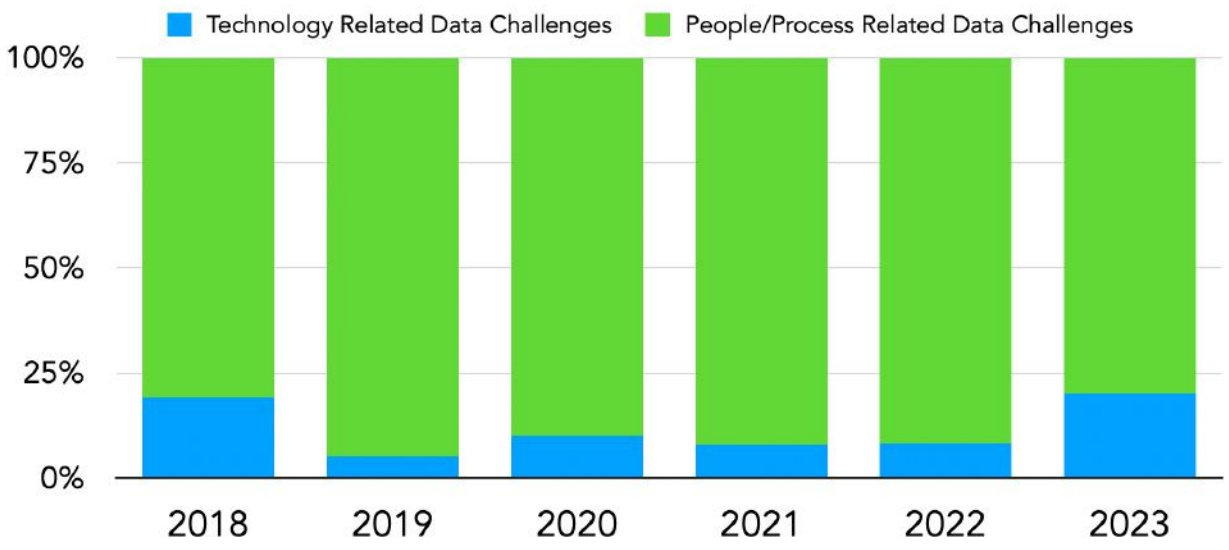
1. Do you have executive positions to support data as an asset?
2. Does the organization track usage of this asset?
3. Are organizational or fiscal controls in place to manage this asset?
4. By and large, are these controls actually executed?
5. Is there general acceptance of the need to manage this asset? (i.e., do people "get it"?)
6. Do serious discussions about this asset feature on the agenda of senior management meetings?

Using this rather obvious set of criteria, it is easy to determine that most organizations are not treating data as an asset but so far we do not survey results on this particular measurement.

Available but not widely known research results

As referenced above, there is a dearth of knowledge about data much less data governance. On that note however, we do have access to two solid lines of research to which I will refer through out this chapter. The first is in the form of the annual (2013-today) data practices surveys conducted by New Vantage Partnerships and are reference able at: https://www.newvantage.com/thoughtleadership. Annually several thousand of the same or similar organizations have been asked the same questions repeatedly providing pictures of how issues are considered over time. Results reproduced here will be referred to as **New Vantage**. A second set of research results come from the collaboration (called the Data Literacy Project) between Accenture and Clique. These results will be referenced as **Data Literacy Project** and are reference-able at: https://thedataliteracyproject.org/. These two efforts have provided a good framework that can be used to dive further into research in this area.

One of the New Vantage results has been: *what percentage of your data challenges are people/process related versus technology challenges?* The consistent answer (see figure) continues to surprise: not once since 2018, has the percentage of technology challenges risen to above 20%. This means that for more than six years, **everyone** should have known that the people/process dimension of DG represents the largest challenge. Yet very little organized research beyond surveys has been conducted into this area.



Consider please, what group in your organization is charged with decreasing the number and impact of people and process-oriented data challenges? This is precisely the role that your DG organization must address in your organization. **If not DG then whom** in your organization is responsible for improving the people and process aspects of your data operations?

It is crucial that DGs provide a wholistic view of minimally the above detail but also include data's role in the organization, how individuals can assist, and where to go for more information.

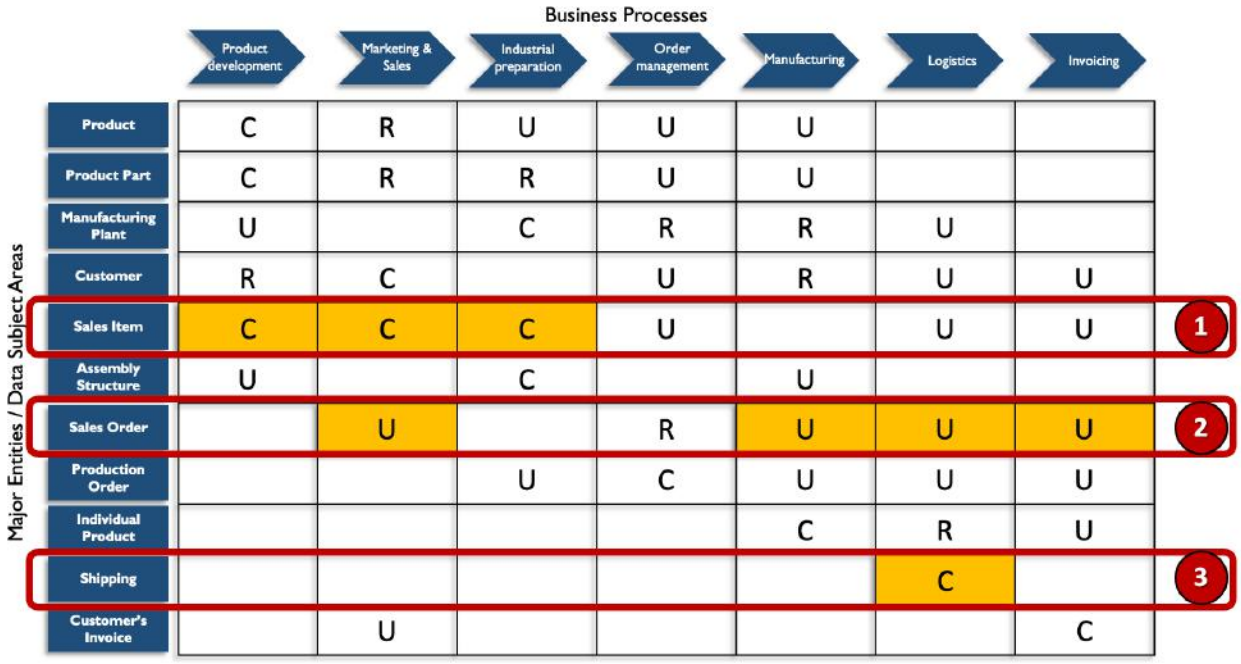
Using data to better to support the organizational mission

This section's title: **using data to better to support the organizational mission** must be the mission of any DG program. But first a specific word about data ownership (bad concept) and Data requirements ownership (good concept).

Avoid a first (and always a major) misstep: trying to assign data 'ownership.' While it is tempting to "establish data owners" as a goal of data governance–it is usually a bad idea. However, many are familiar with the process architecture practice. It correctly embraces and leverages the term "process owner" as the single individual responsible for the integrity of the process design, implementation, and improvement.

While it makes intuitive sense, the concept of data ownership has caused more DG effort to fail than any other. As soon as you allow an under-informed individual (or group) to 'own' any data items, they begin to make decisions about the data that optimize it from their local perspective. If your organization does not formally manage a process architect, skip to the next paragraph. If it does, careful analysis will yield maintainable, high level process/data interaction matrix called a CRUD matrix–showing data/process interaction by access type. (CRUD matrices such as the one illustrated

show business processes and their activity type **C**reating, **R**eading, **U**pdating, and **D**eleting various data items–example also courtesy of Dr. Christopher Bradley).



	Product development	Marketing & Sales	Industrial preparation	Order management	Manufacturing	Logistics	Invoicing	
Product	C	R	U	U	U			
Product Part	C	R	R	U	U			
Manufacturing Plant	U		C	R	R	U		
Customer	R	C		U	R	U	U	
Sales Item	C	C	C	U		U	U	1
Assembly Structure	U		C		U			
Sales Order		U		R	U	U	U	2
Production Order			U	C	U	U	U	
Individual Product					C	R	U	
Shipping						C		3
Customer's Invoice		U					C	

If nothing else, these maintainable metadata collections show the interdependencies: data exist only to be consumed by various business processes and only purpose for a business process to exist is to produce data to be consumed by another business process. If you do not have an organization CRUD matrix hand and need to shut down any data ownership conversations, ask the question: "*To whom does the data that accounting stewards belong?*" Since accounting processes data from across the organization, a case could be made that accounting 'owns' much organizational data.

The reason data ownership is such problematic concept is that data persists across business functions. Ownership would only apply to a specific data processing stage. Instead of asking the question, "who are the data owners?" the statement should be that all data belongs to the organization! At best, ownership could only be limited to specific lifecycle phases.

If the organizational culture requires use of the word ownership, then allow ownership of the **data requirements**! Local expertise should be used to specify the size and shape of the specific data items required to perform organizational functions at various stages of data at it is processed.

The role of DG frameworks

All evidence to-date points to frameworks has being useful as:

- System of ideas for guiding subsequent analyses

- Means of organizing measures, project data, and then assessing progress
- Evaluating priorities for data decision making
- Assessing overall functionality
- Moving towards a determination of ROI[7]

For example, a building construction conceptual framework would incorporate bits of wisdom such as:

- Don't put up walls until foundation inspection is passed
- Put the roof on ASAP so that work can proceed in inclement weather
- Make it each construction phase dependent upon continued funding by passing a series of checkpoints

Much has been written about data governance frameworks. I have seen research proposals that anticipate evaluating one type of framework against another. It is far to early to start to 'type' DG frameworks. Non-standard understanding of terms and data concepts lead to 'results' of the sort that were popular at the start of the CDO movement. (Note: researchers have tried and failed to establish correlations between having a CDO and organizational financial performance–similar specious results can be expected until the entire DG profession matures.)

Use the existing DG frameworks to envision what your program should look like given your organizational needs. 'Try each of them on' conceptually and discuss the suitability of each for your organization. Since no two organizations are alike, each organizational DG program must be custom fitted to the organization rather like getting fitted for a suit. The word "bespoke" well describes the design of DG programs that provide good returns on organizational DG investments.

It is quite useful to view representations of various approaches to DG in the same manner that an architect presents sketches of a future building to prospective funders. The utility of DG frameworks generally stops at this point. There are essentially few types of DG frameworks in popular use. (Note: you can see representations of many these at: https://anythingawesome.com/DataGovernanceFrameworksCollection.html) All subsequent are theme and variations on these. Pay no attention to 'proprietary' methods. The goal is to give you something to compare, contrast and consider when designing the first version of your DG organization. (Note: This first version will evolve to a second and third as the organization; DG practices should mature and evolve over time.)

This is where the concepts of stewardship and fiduciary responsibilities come into play. **Stewardship** in this concept is derived from the definition: a person employed to manage another's property. **Fiduciary** is used to describe the nature of the relationship

[7] Interestingly, ROI means risk of incarceration to most DG professionals.

as involving trust, especially with regard to the relationship between a trustee and a beneficiary. This is accompanied by specific duties.

Related Term Definitions

It is now time to introduce a few terms to show both the evolution/etymology of the term DG and the most useful definition of DG.

Let's start with the term **governance**: "Governance is the process of interactions through the laws, norms, power or language of an organized society over a social system (family, tribe, formal or informal organization, a territory or across territories). It is done by the government of a state, by a market, or by a network. It is the decision-making among the actors involved in a collective problem that leads to the creation, reinforcement, or reproduction of social norms and institutions" (https://en.wikipedia.org/wiki/Governance)

Corporate governance is next. Below are three good definitions highlighting different aspects of this evolving concept.

- "Corporate governance - 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 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 Vishnu, 1997.
- "Corporate governance is about promoting corporate fairness, transparency and accountability" James Wolfensohn, World Bank, President Financial Times, June 1999.

Note that the concept of corporate governance is evolving. Just before the pandemic, Jamie Dimon (then head of Chase), led a group of CEOs to proclaim "Maximizing shareholder value can no longer be a company's main purpose."[8] Similarly, the concept of DG continue to evolve.

Well if corporate governance exists then certainly **IT governance** should be a useful concept? It is and is defined as "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. (https://en.wikipedia.org/wiki/Corporate_governance_of_information_technology)

IT governance frameworks 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. Included are typically foci on:

- Strategic Alignment
- Value Delivery

[8] https://www.marketwatch.com/story/maximizing-shareholder-value-can-no-longer-be-a-companys-main-purpose-business-roundtable-2019-08-19

- Resource Management
- Risk Management
- Performance Measures

IT governance is an established discipline with common vocabulary and understanding among those who participate.[9] Of note is the fact that data practices are not typically included as a topic under IT governance or are lightly treated. This may account for or reflect the current slowly maturing state of DG practices.

Data governance has suffered from both too many definitions and inaccessible (by the business) terminology. However, auditors easily get the concepts. Below are some standard definitions of DG.

- 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

Technically they are all correct but imagine the following scenario. Stepping onto an elevator for a minute-long ride and an executive enters the car. As the doors close the executive turns and says, "I've heard you are working on DG. Can you tell me what it is - I'm confused?" Imagine responding with "DG is the exercise of authority and control over the management of data assets." Do you think the executive would 1) find the answer useful and 2) think well of your ability to communicate this concept?

I think the answer no to both questions. A better response to the executive is: "**DG is about managing data with guidance.**" Short and to the point, this definition incorporates self explanatory motivation. When I provide this information (the definition of DG) to most executives, their first question to me is: "So we have not been managing our data with guidance?" The answer usually is: "Only recently have we been managing our data with guidance." Of course the eternal hope is that the executive will be curious to learn more and present an opportunity to become more data literate. Subsequent conversation topics could include:

[9] https://en.wikipedia.org/wiki/Corporate_governance_of_information_technology

- Why is it generally not a good idea to govern all of your data.
- Why DG will never be complete at our organization.
- Why some decisions that involve data are not considered as such.

The Data Literacy Project reports that four out of five executives surveyed were willing to invest time resources in improving data skillsets. This represents a once in a generation opportunity to reach these executives with good DG education. (Note that anyone offering to improve your organization with DG **training** should be ignored–the process requires **education**, not training.)

A Small concentrated team is preferred over distributed (dissipated) knowledge

The next item to consider is what format DG should take. Remember, asking **everyone** to be responsible for (data, data quality, data governance …) has produced the current state of affairs. Organizations assigning new DG duties to existing personnel have two options: 1) incorporate the new duties along with existing duties or 2) assign these DG duties to full time individuals.

When considering this, it is useful to ask: how long will the need to manage data with guidance exist? The answer turns out the be: **you will need your data program as long as your organization needs to have its finance, HR, and planning operations.** Think about it-in the future: Will more or less data exist? Will data collection modes increase or decrease? Will data be found in fewer or more formats? A solid recommendation is to staff with full-time team members dedicated fully to DG. Data literacy and organizational data practice maturity are generally low. Dedicated personnel with interact with each other more–greatly stimulating their individual learning curves. It also makes tracking DG program costs clearer. It is critical to begin to build organizational DG capabilities. This can best be started with dedicated teams with a clear ROI. Against these, results can be evaluated.

Using data strategically

The next question is **on what** do we focus these DG efforts? In regulated environments, these efforts are often compliance driven. Key is to approach these efforts in the same manner. Do we think that regulations will increase or decrease in the future? If increasing, then it seems useful to 'get good' at implementing compliance driven changes. If nothing else, you may gain an implementation advantage over the competition subject to the same data regulations but perhaps not able to implement as efficiently or effectively. Data regulation compliance can become an valued organizational capability with an easily determined ROI.

Outside of compliance, organizations strive to use data strategically with either efficiency/effectiveness or innovation goals. Personal interaction with more than 1,000 organizations indicates that about ½ have clearly articulated strategic goals and

objective measures supporting goal achievement at the organizational level. Absent these, it is not possible to improve the manner in which data supports this Jell-O strategy. I also find universal distain for 3-5 year plans, most of which fell apart rapidly with the onset of the Covid-19 pandemic. So just a word of caution, check your organizational strategy to ensure it has clear objective and measures before attempting to improve how data can support it.

Strategy is About Why

*…it's not **what** you do, it's **why** you do it…*

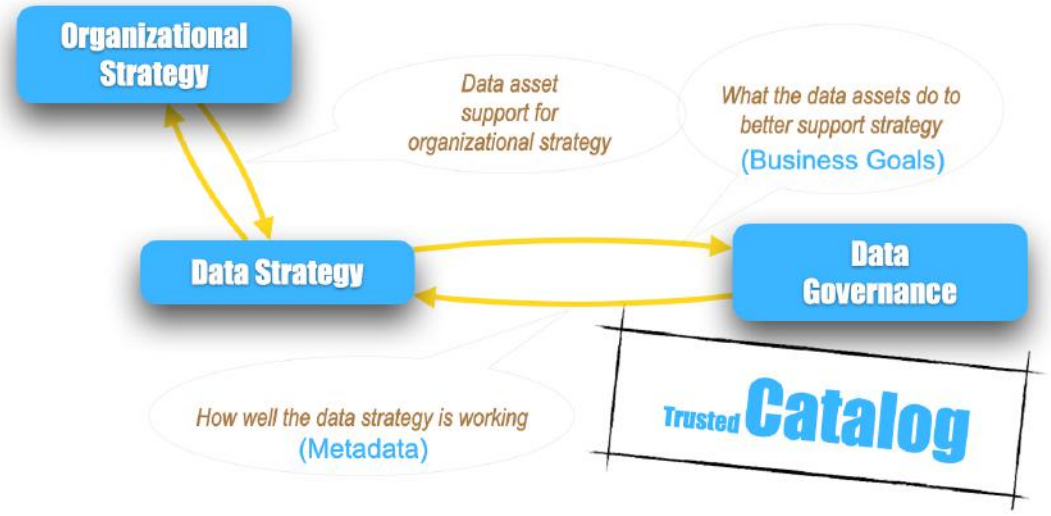
Among many great TED Talks, Simon Sinek's "How Great Leaders Inspire Action" is a favorite. Recorded in 2009, Sinek's talk has enjoyed more than twenty-five million views. His point is quite simple: most of us are very good at describing what we do, and some of us are good at describing how we do things. Not as many of us are good at describing why we do things.

Strategy is the highest-level guidance available to an organization, focusing activities on articulated goal achievement and providing direction and specific guidance when faced with a stream of decisions or uncertainties. More succinctly, strategy is a pattern in a stream of decisions. This pattern must be supported by data or it will not be possible to determine if the strategy is correct or working.

What is data strategy?

Data strategy is the highest level guidance available to an organization, focusing data-related activities on articulated data program goal achievements and providing directional and specific guidance when faced with a stream of decisions or uncertainties about organizational data assets and their application toward business objectives. The data strategy must be understood and supported at the organizational level. Only with this level of scrutiny and involvement can a true systems view be applied to the challenge of improving how data can support strategy.

Working together: data and organizational strategy?



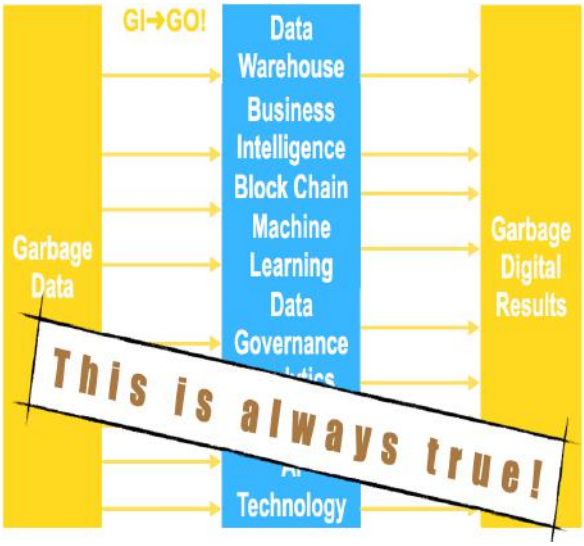
The figure indicate the close relationship among organizational strategy, data strategy, and data governance. Two key aspects of the interaction are: 1) express the data strategy in terms of specific business goals, and 2) ensure that the language of DG is metadata.

Strategic committment: Program versus project focus

A commonly asked question is: when will you be done? This is a warning that the individual considers DG a **project**. Organizations failing to implement DG at the program level (as a program) are unable to view the totality of their data challenges holistically and the solutions fail. Many organizations require a second or increasingly a third DG 'reset.'

Digitizationing

One of the more important areas that DG can be focused to support is 'going digital.' Once again, many vendors have offerings and expertise in these areas. DG sets the standards required to support digitization because, you cannot 'digitize' without a good data capabilities foundation. Garbage in, garbage out is always true. At this point, effective DG is a requirement for digitization, otherwise you can will be unable to trust any digital system outputs.



A watchful eye toward the US Federal Government (FEPA)

Finally on the **what** question (yes–we are still in **what**), it will be useful to observe the progress being made in the US Federal Government. As part of my service as a DoD employee, our group was often sent to 'learn from the private sector.' Now the situation has been reversed. In 2019 the **F**oundations for **E**vidence-Based **P**olicymaking **A**ct was signed into law. Three specific aspects of the law make this especially interesting to aspect of DG to follow. They are:

- Explicitly non-political CDOs must be established separate from CIO roles. From a DG perspective, organizations have been slower to adopt CDOs with non-CIO reporting role.
- Government data is now open by default, and must maintained using open standards. In just a few years, the Federal agencies will have developed a great deal of expertise in these areas.
- Use of open data and open models is required in policy evolution. Policy changes are only permitted with both models and datasets specified prior to the analyses and decisions.

Collectively these efforts, if fully implemented, will improve governmental decision-making and overall effectiveness. More importantly, all impacted Federal organizations are also rapidly, developing and implementing DG as compliance activities still further increasing the pool of DG professionals world-wide.

Breaking through the Barriers of Data Governance

There are a host of barriers to implementing DG. This include the usual failures to include change management and cultural refocusing as key dependencies. While the accounting profession has had literally millennia to develop GAAP, no such guidance exists for DG. There is a vast tendency to depend on technologies that are incapable of acting as silver bullets.

An example of these difficulties was illustrated in 2020 when Forbes ran an article on airline valuations.[10] It purported to show how the airlines were monetizing the data in their frequent flyer programs. However, the buried lede was that in 2020, both United and American Airlines were valued at 10s of billions of dollars less than the anticipated value of the data in these programs. You had better believe that if airline leadership could have unlocked that value during the time most were avoiding flying (the pandemic), they would have unlocked it asap! The fact that they were unable to do so highlights the uphill climb that poorly fitting DG efforts face.

Some basic DG execution principles follow:

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

- Ensure that the organization's data strategy in properly aligned with the business strategy. Implement a regular processes with key stakeholders to ensure proper alignment.
- Ensure that data debt is properly being managed and the process is under statistical control.
- Perform a capability maturity assessment or 'reassessment' to determine the required maturity. If the maturity levels are not meeting expectations, ensure that there is a remediation plan with a properly monitored work arounds.
- Consider refresher training for your knowledge workers and data professionals, e.g., Data Stewards, Architects and Engineers as a feedback mechanism for determining needed improvements and remediations.

Based on the organization's strategy, the DG group must determine are they to initially follow a model primarily focused as a:

- Utility–back office, efficiency goal
- Steward–more asset focused, quality goal
- Enabler-strategic partner, innovation goal

This should be determined through the building of the data strategy. If an organization striving toward a modernization transformation, DG should trend to an 'enabler.' To measure the effectiveness of an enabler, DG standards should be repeatable and statistically stable. The focus can be changed at a later stage but can usefully focus effort and discussions during initial phases.

Hopefully your organization will be spared major data catastrophes. It is more likely you will experience one or more in the future. In this event, attempt to learn as much as possible from the event. Take for example, the story of two major banks in the process of consummating an arranged marriage. The deal came down to a single spreadsheet containing many rows, each representing an asset. If an asset on the spreadsheet was to not be transfered, that row was hidden with agreement by both parties. After final agreement was reached, the spreadsheet was handed to a junior associate who was told to 'make is look nice for the Judge tomorrow." Unfortunately late in the evening junior accidentally unhid hundreds of rows and did not notice! Presented to the Judge as the golden copy, the judge would not reverse-even on appeal.[11] As you might imagine, DG practices around the use of spreadsheets is quite extensive. I assisted one organization with the elimination of more than 400,000 legacy systems of a certain type. The list of preventable spending continues.

Unfortunately, the conversations have been generally unsatisfactory. Key to getting started with **data valuation** is to add up 'at least' instead of attempting to master the entire costs. I justified an investment into an organizational repository at one

[11] https://www.businessinsider.com/2008/10/barclays-excel-error-results-in-lehman-chaos

organization with a business case built on the premise of saving everyone in IT **1 hour annually**. The organization conducted surveys asking if the one hour savings was achieved. It was!

When determining the internal and external value of data two prerequisites exist: first, business and data strategies must support data monetization and second, DG must be effective and properly measured. Components of data value can include:

Internal

- Properly managed data debt
- Efficient usage of cataloging and master data management
- Hight trust in supplier and customer data integration
- Measured positive ROI

External

- Organizational data monetized in a public market or exchange
- Organizational data becomes a profit center
- Organizational data becomes a band-aid of adhesive strips

Sometimes it is easier to highlight the value with unfortunate examples with clear costs to society. Early COVID monitoring was inhibited because health care workers did not know to save MS Excel data sheet and workbooks as .xlsx instead of .xls files. The difference, unknown to the users, was that the older .xls files dropped all rows beyond the 16,000th or so row without warning. We will likely never know how much better performing the early monitoring systems were because all the errors are in one direction.

On a more cheery note, an agency charged with home evaluation/intervention, discovered that 40 questions on its evaluation assessment were immaterial. This shortened each interview by half and ultimately shifted more than $1 million from overhead to service delivery.

In terms of execution, DG should be viewed as an iterative process that the organization is striving to **get better at!** Each cycle focuses on aspects of the various data challenges with a goal of eliminating or reducing the impact of a specific constraint. To understand the important of this shift in thinking about DG, consider the circumstances where a plan was the goal. It was former President and General Eisenhower who said:

> "*In preparing for battle I have always found that plans are useless, but planning is indispensable*"[12]

Mike Tyson's version is that *everyone has a plan until they get punched in the face*. A team knows how to react to unforeseen challenges and efficiently address the ones they have planned for. The PDCA cycle provides operational context.

[12] https://quoteinvestigator.com/2017/11/18/planning/

Chapter Summary

The word **bespoke** has evolved from a verb meaning 'to speak for something', to its contemporary usage as an adjective. Originally, the adjective bespoke described tailor-made suits and shoes. Later, it described anything commissioned to a particular specification. Wikipedia

The difference between data analysis capabilities and data requiring analysis is increasing. DG will continue as a maturing and growing field and can only be assisted by increased research into the various challenges outlined. Practice standardization and improvement are clearly the next steps on this industry's maturity curve. As a new discipline, DG works best directly addressing the manner in which data is used to support achievement of organizations strategy. There is no one best way and right now there isn't agreement on terminology, hence anything. Consequently, the only way to obtain a positive ROI on investments in DG is to ensure that **your** data is successfully leveraged using methods (**your** data strategy) that **your** knowledge workers and **your** executives understand.

The goal is to improve DG effectiveness and efficiencies (and the data itself) over time. The more data literate the organization, the easier the transformation. Perhaps now is more understood about the phrase quoted at the beginning of the chapter:

This database ain't big enough for the two of us

– Bumpersticker seen on an automobile in Texas

Acknowledgement

My colleague Rob Greaves made many helpful suggestions that were incorporated into this chapter.

I appreciate Dr. Christopher Bradley granting me use of the two concepts note herein.