

Uma Gupta • San Cannon



A PRACTITIONER'S GUIDE^{TO} DATA GOVERNANCE

A Case-Based Approach

Uma Gupta • San Cannon



**A PRACTITIONER'S
GUIDE^{TO} DATA
GOVERNANCE**

A Case-Based Approach

A PRACTITIONER'S GUIDE TO DATA GOVERNANCE

A Case-based Approach

UMA GUPTA

*Director of Business
Analytics, University
of South Carolina at
Upstate*

AND

SAN CANNON

*Associate Vice
Provost for Data
Governance and
Chief Data Officer,
University of
Rochester*



United Kingdom – North America – Japan – India
Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2020

Copyright © 2020 Emerald Publishing Limited

Reprints and permissions service

Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-78973-570-3 (Print)

ISBN: 978-1-78973-567-3 (Online)

ISBN: 978-1-78973-569-7 (Epub)



ISOQAR
REGISTERED

Certificate Number 1985
ISO 14001

ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.



INVESTOR IN PEOPLE

Contents

1. Foundations of Data Governance
2. Impact of Organizational Culture and the Need for Change Management
3. Communication: Key to Success
4. Data Strategy
5. Data Governance Frameworks
6. Data Governance Components: Data Quality, Literacy, and Ethics
7. Data Governance Maturity Models
8. Summary Case Studies
9. Detailed Case Study
10. Execution Roadmap

Notes

Resources and References

Index

1

Foundations of Data Governance

Introduction

Data governance looks deceptively simple on paper. It is not. It is difficult. It is complex. It is often abstract and ambiguous. It is a long road where leaders must remain committed to continuous improvement even as they face several roadblocks to achieving their goals. It is a game of persistence and creativity where those who understand the power and potential of data work diligently to tame the ocean of data in their organizations with the ultimate goal of leveraging data to achieve organizational excellence, market power, and exceptional customer service.

There are many hundreds of books and white papers, journal and magazine articles, and internet resources on data governance. Research and theories are essential to advance an emerging, interdisciplinary, and fast-moving field like data governance. Many of these resources are vital to helping researchers and practitioners advance the field of data management and governance. Leaders need education and training on a diverse set of strategies to implement data governance in their specific industry since each industry has its own data challenges and regulatory and compliance requirements. Frameworks are necessary to help organizations implement data governance, particularly those that are just embarking on this journey. A rich and diverse set of knowledge resources and outlets are absolutely essential to the advancement of any field.

However, mainstream resources on data governance can sometimes lead one to believe that there is a simple formula to achieve overnight success. How-to articles may give the impression that implementing data governance is nothing more than following a set of logical steps as it relates to people, processes, and technology. But data governance practitioners know better. They face a unique

set of challenges in the field that extend beyond the immediate scope of cutting-edge research and theoretical frameworks. Those who are in the data governance arena know all too well that success is slippery at best, and that the road to achieving optimal governance through excellence in data governance execution will forever remain an ongoing challenge.

We recognize that while theoretical knowledge is essential and foundational, it does not always neatly translate into success in the real world. Our focus, therefore, is to help the data governance practitioner in the field with tools and tips to navigate the unique challenges that he or she faces. Our goal is to leverage and utilize the rich foundation of data governance research, strategies, tools, techniques, and frameworks that exist today to deliver a portfolio of practical ideas that can assist data governance practitioners and leaders to create and implement successful data governance initiatives and programs in their organizations.

Why Invest in Data Governance?

It may be tempting to treat data governance as another passing fad. After all, the history of business is filled with fads. The reason why data governance cannot be sidelined as a fad is that regardless of the business unit or the industry or the geographical location of an organization, every company, institution, and society rely on data to make progress. Today, data have moved from a backdoor operation to be the central nervous system of an organization. Relevant and timely data dictate the growth, survival, success, and sustainability of an organization. We know that organizations manage and monitor their financial resources with care and attention. Since data are the drivers of revenue, even more care and attention need to be paid to data and their integrity and management. It is a dangerous idea to treat data as a secondary resource that does not deserve its own dedicated and knowledgeable team. Furthermore, today there is evidence showing a direct link between how well an organization manages its data resources and its financial performance. This should come as no surprise because data are the back belly of effective decision-making, which, in turn, drives financial performance.

Finally, employees need data to make good decisions. When employees don't have the right data at the right time or they cannot find the data or have the wrong data, the quality of decisions is affected, and this can have a reverberating effect on the entire organization. It leads to employee frustration and loss of

productivity. For this reason alone, if nothing else, data governance is here to stay. And organizations that embrace data governance will benefit and leapfrog those who ignore the value of data as a resource.

What Is Data Governance?

There is no universal definition of data governance. Why? Because data governance, by its very nature, cannot be defined in isolation. It is like a tailored suit that must fit the person. Like creating a bespoke suit, designing data governance for an organization takes time and effort. It may make sense to start with something off the rack that sort of fits and tailor it to suit the wearer. But even suits on the rack have variations in color, fabric, and style. The suit that is chosen depends on the wearer and the occasion. Similarly, the creation of a data governance program depends on the organization: programs and initiatives must align with its unique cultural characteristics and data needs. In other words, the validity of any definition of data governance is intimately and intricately tied to a variety of organizational characteristics such as the organization's vision, values, and strategic priorities; portfolio of risks; people and culture; organizational structure; stakeholder demands; regulatory and compliance requirements; data management guidelines and protocols; maturity stage of data governance within the organization; and market dynamics, to name a few. These factors make it essential for data governance leaders and senior management to develop a tailored definition of data governance. We, therefore, recommend that every organization carefully craft its own definition of data governance that takes into account all critical and relevant factors to ensure stakeholder engagement and buy-in. The process may start with an “off the rack” definition but it should not end there. Without a customized definition of data governance, it may be difficult to ensure the success and sustainability of data governance strategies and initiatives in an organization.

How an organization defines data governance will determine how successful it is in creating and executing its data governance agenda over the short and long term. Why? Because data governance is the responsibility of the entire organization, not just the C-suite or the IT department. Although senior leadership is largely responsible for setting the tone for how data and their usage are valued in an organization, a clear, concise, and customized definition of data governance brings the message about the value and power of data to all employees. Every employee who relies on data for performing his or her

professional tasks directly or indirectly contributes to the data governance pie, whether it is a small slice or a significant one, or a formal or informal role in data governance initiatives and projects. Every employee therefore has a professional obligation to ensure that data governance programs and initiatives in the organization succeed. Hence, defining data governance within the organizational context should not be taken lightly nor should it be rushed as it plays a crucial role in winning employee buy-in and commitment. We, therefore, recommend that all data governance initiatives begin with a customized definition of data governance. We also recommend that organizations periodically review their definition of data governance to ensure that it remains strategic and current given that all organizations are living organisms.

What are the critical elements that should go into a customized definition of data governance? This question is best answered by looking at all the key factors that impact and influence data as an asset. As we know, there are many dimensions to leveraging data for the long-term success of the organization. These include data accessibility, quality, consistency, and integrity; roles and responsibilities of individuals with access to data, policies, and procedures that govern who has access to what data, when, and why; and processes that mandate how data are to be used in an organization. Not all data that reside in an organization are the same type. Given that there are many types of data (master data, big data, and reference data, to name a few), the definition of data governance must be flexible to include all types of organizational data.

We present below several popular definitions of data governance as a starting point for developing a customized definition that takes into account the elements we mentioned before. Consider these some of the suits on the rack:

1. Data Governance is a collection of practices and processes which help to ensure the formal management of data assets within an organization. (DATAVERSITY)
2. Data governance is defined as the exercise of authority and control (e.g., planning, monitoring, and enforcement) over the management of data assets. (DAMA DMBOK)
3. Data Governance is a discipline that provides clear-cut policies; procedures; standards; roles; responsibilities; and accountabilities to ensure that data is well-managed as an enterprise resource. (DGPO)

4. Data governance is a data management concept concerning the capability that enables an organization to ensure that high data quality exists throughout the complete life cycle of the data. The key focus areas of data governance include availability, usability, consistency, data integrity and data security and includes establishing processes to ensure effective data management throughout the enterprise such as accountability for the adverse effects of poor data quality and ensuring that the data which an enterprise has can be used by the entire organization. ([Wikipedia](#))
5. Data Management International (DAMA) defines data governance as “the exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets. Data Governance is high-level planning and control over data management.”
6. ISO defines data management as “the activities of defining, creating, storing, maintaining and providing access to data and associated processes in one or more information systems.”

Key terms in these definitions include practices, procedures, standards and processes, roles and responsibilities, data consistency, integrity and management, formal management, data assets and enterprise resource, authority, and accountability and control.

By integrating the above foundational terms of data governance with the unique characteristics and needs of data in a given organization, such as accessibility, quality, consistency, integrity, maturity, and compliance requirements, the beginnings of a customized definition of data governance start to emerge.

The challenge facing data governance leaders is to make standard definitions of data governance come alive within their organizational context with the sole purpose of winning the support and commitment of all stakeholders. The purpose of this exercise is not to create a technical definition of data governance that only those who are intimately involved with data can understand or accept. The goal is to define data governance in such a way that all users of data will understand the full implications of data governance and be willing to commit to doing their part to protect the data assets of an organization, which is a precious resource.ⁱ

Why Is Data Governance Difficult?

There are many reasons why data governance is a daunting task for organizations, governments, and societies. Achieving a reasonable level of data governance sophistication and maturity is a long road with many ups and downs. The final destination is one that is continuously shifting and emerging. Failures are common, even though one rarely hears about it in the news or in popular trade magazines.

Data Control Is Not Easy

First, at its core, data governance is about understanding and leveraging one of the most valuable assets of an organization, namely, data. This is not as easy or simple as it sounds because there are many moving parts to every organization. This includes establishing strategic priorities as it relates to data, their integrity, usage, and accessibility; creating a bold, yet flexible, baseline of data management frameworks and systems; continuously scanning and identifying opportunities in the marketplace by leveraging the data the company owns; monetizing the data either through upward or downward channels or both; and ensuring that there is a strong bridge between the data assets the company owns and decision-making at all levels within the organization, to name a few.

Data Problems Are Usually Business Problems

The second reason why data governance is difficult is because what appears as data-related problems are rarely that. Instead, evidence shows that most data problems are usually business problems. Data problems are like a leaky pipe – something is broken somewhere in the business and the source of the leak must be identified and fixed. Data problems cannot be ignored, although organizations do this frequently because they lead to increased cost and decreased revenue. It can lead to employee frustration and loss of productivity. Organizations miss opportunities to enter new markets and provide exceptional service to customers. When business problems are couched as data problems, IT cannot resolve the issue – only the business units can fix it. In organizations with large troves of data, fixing the underlying business problems behind data problems is no small challenge and often politics and turf wars get in the way.

Requires Hard and Soft Skills

Third, data governance demands technical knowledge and exceptional people skills. Part of implementing data governance processes and procedures is structured and logical, but a significant part of successful data governance initiatives requires intuition, creativity, and soft skills. It entails motivating people to do the right thing. It requires implementing best practices in change management and redirecting the ship if things are not going as planned. It requires gaining long-term support from sponsors and senior management and making sure that they remain committed to the cause. Data governance leaders must relentlessly pitch the importance of the work that they do in order to acquire necessary resources since governing data across the organizational spectrum requires time, talent, and money. The entire data governance team must be strong in project management and understand the subtle nuances of an organization's culture without which even the best data governance initiatives are doomed to fail.

Current Report Card Is Not Good

The fourth reason why data governance remains a challenge is because of the current state of affairs. The overall report card on data governance as it stands today is not good. Many organizations receive a low grade when it comes to data governance. Cross-industry studies show that on average, less than one in two organizations use their structured data to make decisions, and that a shocking less than 1% of unstructured data is even accessed, analyzed, or used to make decisions. In many organizations, 7 out of 10 employees have access to data that they should not. Finally, 80% of an analyst's time is spent simply locating the data and formatting the data for further analysis and use.ⁱⁱ Unfortunately, the foundation to build a strong data governance culture is not there today. In many organizations, especially decentralized ones, employees feel empowered to work locally and often strive to do the best for their team, unit, or department. These local maximization efforts often fall short of the global or enterprise-wide optimum that good data governance can provide. This makes the haul longer and more challenging.

Data Governance Is Complex

The fifth reason why data governance is hard is that it is not about solving one problem. It is about solving an army of problems, each with many tentacles all of which have one common source: data. In short, data governance is not just complicated – it is complex. While most organizations understand the importance of data as a valuable resource, executing this vision and value is a daunting task because the nature and scope of data governance problems are highly dynamic and not always easy to spot. For example, how does an organization ensure that the large volumes of data it generates and owns are put to good use instead of languishing somewhere? How can organizations ensure that employee productivity is not impacted by hours lost in searching for relevant data? How can an organization ensure that the quality of its data meets its own internal standards? How can organizations prevent data neglect, in whatever form it may be? How does an organization ensure that every employee understands their role in managing data as an asset? In other words, how does an organization ensure that its data are “fit for purpose,” timely, relevant, and useful?

Data Governance Requires Change

Finally, data governance is hard because it requires change and change is hard. Once an organization identifies the business problems, understands the complexity of the issues, acknowledges the problem with the current state, and trains or hires the right skills, people still have to do something differently than what is currently being done. As we will argue here and in future chapters, not understanding or appreciating the effect that change has on the workforce dooms many projects from the outset. Often, senior leaders assume that employees will readily sign on to whatever change initiative is underway because they see the same value to the organization as leadership does. Not only is that rarely the case, staff affected by change are often only focused on what that change means for them, and it is often equated with a sense of loss. No matter how much better, cheaper, or faster a new platform or process will be, it will require staff to stop doing something they are familiar with or good at and start doing something new with which they are unlikely to have expertise. If not addressed correctly, this can lead to roadblocks from passive resistance to employee turnover.

The Three Core Elements of Data Governance

Those who lead and spearhead data governance know all too well that given the complexity and the long-term nature of such initiatives, it is important to manage expectations, particularly of the senior leadership team. Senior leaders may assume that the data governance leader will quickly fix the broken pieces of data management and data stewardship and develop a strong foundation where the right data are available at the right time from the right source and that they are shared appropriately throughout the organization – in other words, achieve an ideal state with reasonable effort. However, many data governance experts will emphatically state that this is not realistic. Experienced data governance leaders reiterate that data governance is the ongoing responsibility of every key stakeholder. The most challenging task confronting leaders in data governance is to convince each employee in the organization that data governance is their responsibility and not that of IT or a third party.

The value and critical nature of data governance seem obvious, doesn't it? So why is this a big challenge? The truth is most people have some basic understanding of the importance of data and the necessity of managing and protecting them, but they also have their own jobs and responsibilities. The governance of data is not foremost in their minds at work. They may understand and even support data governance on good days but translating it into action can be challenging. The data governance leader or manager, on the other hand, as someone charged with protecting the data assets of an organization, may be lulled into thinking or even believing that most employees are on board with data governance priorities. After all, everyone nods their head in agreement and in unison in meetings about the critical nature and value of data. The truth is employees often agree with the data governance priorities, focus areas, mission, vision, goals, and objectives, so long as it doesn't entail any additional work or change in their behavior. In other words, employees are on board, conceptually speaking, with the need for data governance so long as it is about what *other employees* must do to modify their behavior.

The core elements of data governance often cited in popular literature are people, processes, and technology. This is true, but this is only a partial list. The success of data governance initiatives depends on three additional and equally important core elements that are rarely mentioned. These are communications, culture, and change management.

Communications

Communications is the key element to success in data governance. Often, people even within an organization do not speak the same data language. The dialects and nuances of words and meanings as it relates to data, information, knowledge, and decision-making are as different as the people in an organization. It often takes years of effort and practice before everyone in an organization uses a common language when it comes to data and their concepts, elements, and usage. Further, leaders know that a communication style that works in one organization may or may not work in another. Seasoned data governance leaders recognize that communication strategies that worked in one environment or organization cannot always be easily transferred to another environment and even then, do not guarantee the same results. In other words, there is no proven communication manual or standard formula on how to build a fail-proof data governance initiative. Communication strategies, which underpin the success of data governance initiatives, must be customized and tweaked not only for every organization but also for each unit within a given organization. We have dedicated all of Chapter 3 to effective communication strategies for data governance.

Culture

The second and third core elements for data governance success are a strong understanding and sensitivity to the data culture of the organization coupled with meaningful change management strategies. We believe that without a deep insight into an organization's culture, change management strategies cannot be triggered and without both (culture and change management), many data governance initiatives are bound to fail.

Those with strong technical expertise may sometimes underestimate the impact of an organization's data culture on data governance and its influence on the long-term success and sustainability of treating data as a precious resource. They may be tempted to focus primarily on two of the three core foundational elements of data governance, namely, processes and technology, and ignore or give secondary importance to people, the third element. This greatly increases the chances of failure of even well-crafted data governance initiatives. In fact, if one were to analyze how to gain meaningful and long-term stakeholder engagement with data governance based on real-world success stories, the answer is obvious: it would be through a deep understanding of and sensitivity to the organization's culture and its propensity for change.

Understanding the importance of something doesn't mean it is easy to do. Culture cannot be imported or imposed. Culture evolves and changes right under one's nose. Culture is different from team to team, leader to leader, unit to unit, function to function, and location to location. It is complex, nonlinear, emerging, hidden, and immeasurable. And yet, culture is at the very core of how a company does business, how its employees engage with one another and its customers, and how they create, use, manage, and share data for the well-being of the organization and the common good. The unspoken, but well-known, truth is that not everyone in the organizations understands, appreciates, or is committed to data governance. Many believe that data governance is an IT issue and hence are eager to prove that data governance is overrated and frankly, a waste of *their* time and money. Culture plays an important role in marginalizing employees who fail to understand the importance of data to the health and well-being of an organization.

Change Management

Culture and change management go together, although each demands its own detailed planning and execution along with a unique portfolio of framework and tools. Change management is not unique to data governance. In fact, it is something that is often overlooked as part of a data governance program. If organizations think about change management at all, it is most likely in the context of a technology project where a new platform or application will obviously require new processes and training. Yet change management is more than technology. It can be defined as “the discipline that guides how we prepare, equip and support individuals to successfully adopt change in order to drive organizational success and outcomes.”ⁱⁱⁱ There are a variety of models to help organizations address change management, and we will discuss a few shortly, but the first step to successfully implementing change is acknowledging that change is difficult and needs to be specifically addressed. Because of its importance to data governance success, we have dedicated the next chapter (Chapter 2) to Culture and Change Management.

Who Is Responsible for Data Governance?

In the above sections, we have emphasized the fact that data governance is

important for the success and survival of an organization. In fact, without data governance, organizations can face legal penalties, incur severe financial losses, lose customer trust, foster internal conflicts, and suffer reputational damage to their brand, to name a few. On the positive side, there is a direct correlation between financial performance and effective data management. Companies, where a senior leader is held responsible for data management, fare better financially compared to their peers. Employees in such organizations have a greater probability of locating the right data at the right time and leverage data access and usage to make better decisions. Leaders in such organizations can make the right strategic market moves because they have an in-depth understanding of the current and future needs of their customers. Such organizations can wow customers with creative and innovative high-quality products and services. Regulatory and compliance violations are negligible or nonexistent in such environments. In short, the benefits of data governance make data governance a core component of an organization's strategy.

Given the importance of data governance to the health of an organization, who should be responsible for data governance? The answer is simple: everyone. Everyone in an organization who has anything to do with data is responsible for data governance. Anyone who creates, moves graphs, stores, documents, reports, shares, or generally thinks about data has a role in data governance. Some roles have special names and responsibilities which may vary from one individual to the next, from one department to the next, and from one function to the next. However, everyone and anyone who has anything to do with the organization's data has some level of responsibility for data governance.

We want to shed light on the common misperception that since data storage and security issues often reside with the IT department, data governance is an IT responsibility. This is not true. In Chapter 5, we outline the relationship and differences between IT governance and data governance. We emphasize the fact that it is the business side, not the IT side, that is solely responsible for the creation, processing, usage, sharing, and life of the data. The responsibility of IT is to ensure that the data are stored and secured using optimal infrastructure. It is the responsibility of the business user to know how the data were created, why they exist, where to find them, how and when to use them, how they interact with other datasets, who has rights and authority over data, and information about protocols regarding sharing data. This is the responsibility of each user who uses a specific set of data, regardless of their level or functional expertise, in an organization. It is *not* the responsibility of the IT department.

In summary, data governance is a challenging endeavor. Ignoring it only makes the problem more monstrous and unmanageable. This book presents a strategic, systematic, and structured approach to addressing this challenge. We rely on our GPS to get us to our destination. Consider data governance as the GPS that guides an organization to achieve its strategic goals of capturing the underlying power of data to facilitate timely and effective decision-making.

Goals of This Book

This is a how-to book that recognizes that there is no perfect or complete how-to book! It is a field manual that recognizes that the field of data governance is dynamic and ever-changing. We wanted to write a book that helps practitioners be better prepared to deliver their mission, goals, and objectives by managing expectations, understanding the organization's culture, and achieving small, yet meaningful, successes in order to have a lasting influence on the organization. We achieve this through situation analysis, case studies, interviews with experts in data governance, and by sharing meaningful and timely research and capturing best practices.

This book offers guidance and practical tips to readers in the following areas:

- Understand the data governance ecosystem and its elements;
- Customize the definition of data governance to align with organizational goals and priorities;
- Assess organizational culture – power vs influence;
- Assess the current state of data governance affairs – going beyond the maturity model;
- Implement exceptional, yet flexible, communication strategies;
- Provide practical and sustainable ideas for data governance strategy;
- Address emerging data governance issues and challenges;
- Benefit from lessons from leaders in the field.

We recognize that there are many challenges and pitfalls in advancing the

governance goals and objectives. We understand the difficulties and challenges that data governance leaders face every day to advance their organization's ability to handle and leverage data. We don't have all the answers, but we wrote this book to integrate theory and practice into meaningful and practical tips to help and guide leaders and novices to navigate the big landmines in data governance with creativity and astuteness. We offer practical and hands-on advice, using case studies and real-world experiences, on how to build and sustain a strong data governance initiative that is business-centric with full engagement of key stakeholders.

Questions to Consider

What Do You Mean by Data?

One of the first questions that any data governance program should address is the definition of data. This is more than a scoping exercise: not all data will need to be governed or be governed with the same approach. As we will discuss further in Chapter 3, it is important to understand what is meant by data in order for a data governance program to be effective. For many people, the default assumption is that data are numeric representations of business concepts that fit nicely into rows and columns of a spreadsheet with variable names or concept labels as the columns and the individual data elements in the rows. For many data governance efforts, especially in the early stages, this is a perfectly reasonable definition and a good place to start. Many organizations aim to formalize data governance to reign in the multitude of spreadsheets and other tools that drive the daily business of the enterprise. Therefore, data should be defined keeping in mind the broad scope, reach and boundaries of the data governance program.

The boundaries must be clearly defined in order for any data governance work to be effective. For some people, data are not just numbers in a spreadsheet but also include text, video, audio, or in some cases any digital content from which information can be extracted. Such an all-encompassing view is at odds with the more limited “it looks like a spreadsheet” definition and that tension will torpedo any data governance effort if different participants have different definitions. For the “everything digital” crowd, documents and email are data that need to be governed as

part of a data governance program. For the “looks like a spreadsheet” crowd, such elements are out of scope and should be dealt with separately so the focus can be on handling “real” data.

To be very clear: there is no correct answer. There needs to be an agreed-upon definition at the outset of any data governance work in order to be able to effectively use any of the advice and strategies suggested in this book. Are data what go into analytics? Or does the output from analytics also count as data? What formats are considered data for the sake of data governance? The definition that an organization starts out with can certainly change over time as data uses and business practices change and as data management and data governance mature. It is imperative that the definition be agreed upon and clear at every stage, regardless of what the definition is.

1. What data governance processes or practices are in place in your organization? What successes have been achieved? What roadblocks are holding back progress?
2. What are the data to be governed?
3. What are you hoping to learn from this book? How will you use that information?

2

Impact of Organizational Culture and the Need for Change Management

Introduction

In the last chapter, we emphasized the importance and impact of organizational culture on data governance. The success of data governance initiatives is intricately tied to organizational culture and its philosophy of change management. Organizational culture, from a layman's perspective, is probably one of the hardest things to describe because although it has several visible elements that employees can point to, it also has many invisible, yet highly impactful, elements. Employees at many organizations have a difficult time clearly articulating their organization's culture, and even when they do, it is often through anecdotes and not direct descriptions.

In spite of the challenges we all face to articulate and define organizational culture, the visible and invisible elements of culture together is one of the key factors that determines the level of performance and engagement of an employee with an organization's mission and goals (Cameron & Quinn, 2011). Organizational culture is ever evolving and, from an employee's perspective, is highly impactful, personal, and subjective. Organizational meeting agendas rarely include or encourage discussions about organizational culture, its nuances, and impact on employee morale and performance. Yet, intuitively we all know that culture shapes behavior. When an employee accepts a role in a different organization, the individual quickly modifies his or her behavior to adapt to the culture of the new organization. In this case, the definition of “standard organizational behavior” shifts for the employee. As individuals, we intentionally or unintentionally judge every norm in the organization as good or

bad. We store the cultural similarities and differences in memory and modify our behavior accordingly to adjust to the expectations of our new employer. Something as simple as work hours can prove a useful example. Some organizations value early effort and there may be a rule or expectation, written or otherwise, that employees clock in by a certain time or prior to the boss. Other organizations may have a different time expectation and it may be seen that leaving the office before 6 or 7 p.m. indicates lack of dedication regardless of the start time. If an employee moves from the former to the latter, they quickly internalize the new cultural norm and adapt their schedule to meet it. The same is true for all aspects of organizational culture; we adapt and change our professional behavior where it is absolutely critical to fit into the current organizational culture and retain behavioral traits from previous cultural exposures and experiences, even though these hidden beliefs, norms, and cultural traits may be invisible to our colleagues.

In many surveys over the years, senior leaders state that the importance of culture precedes that of strategy and even the organization's revenue-generating business model. Given the critical role that culture plays in the overall success and sustainability of an organization, it is only natural that culture will impact how an organization governs its data. It is therefore prudent for advocates and leaders of data governance to step back and gain a deeper understanding of the culture of their organization. The truth of the matter is that the best of technologies, strategies, management tools, innovative approaches, and lofty goals are doomed to fail without a culture that supports and sustains progress. People can work together toward a common cause or mission if and only if their behavior aligns with the values, beliefs, and cultural and social norms of the group to which they belong. Hence, understanding organizational culture and its strengths and weaknesses must be the first step in the ladder to governance.

Organizational Culture

Schein (2010) defines organizational culture as “a pattern of shared basic assumptions learned by a group as it solved its problems of external adaptation and internal integration, which has worked well enough to be considered valid, and therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems.”

Culture includes, according to Schein, observable patterns of behavior as to how people interact with each other, group norms, espoused values, written and

unwritten rules, policies and principles, organizational social climate, skills and competencies that are valued and nourished, mental models and cognitive biases and frameworks, decision-making styles and preferences, and shared meanings. This is a long list of influencers on organizational culture and even if one cannot dedicate time and effort to studying each influencer and its impact on the organization, it is essential to at least be aware that these forces will impact your success as a data governance leader, advocate, or activist. While we recognize, based on the work of leaders in organizational culture like Schein, that culture is an abstraction, overlooking culture is a guaranteed way to derail one's own success.

When employees have a clear understanding of the culture of their workplace and can articulate it in clear terms, we say that the organization has a strong culture. This does not necessarily mean that the culture is good or progressive or enabling, it simply means that the culture is well understood. In such cases, there is a high level of uniformity in how employees across the organization regardless of the nature and scope of their job or hierarchy define and describe the values, beliefs, and cultural norms, how they conduct themselves, and how they make decisions. A weak culture, on the other hand, is one that employees struggle to define and articulate. This often happens when there is a cluster of different cultures within the same organization. The roots of breeding of weak cultures can be traced to top leadership that does not expect or demand alignment with the fundamental tenets of the organization or, even worse, has not clearly articulated the cultural doctrines and beliefs.

From a data governance perspective, it is worth noting that organizational culture has been shown to influence the ability of an organization to grow and to reach its full potential. Studies show that culture affects organizational performance, competitive advantage, and its ability and effectiveness in responding to external threats. Culture influences the creation, usage, and sharing of knowledge. It impacts the types of knowledge that is created and the quality, quantity, and timing (sporadic or continuous) of knowledge creation and sharing. Culture influences the value that an organization places on knowledge creation and how it is leveraged and shared. It determines if an organization rewards those who create and share knowledge. It shapes how employees interact with each other and with internal and external stakeholders. Finally, organizational culture determines if knowledge is valued and monitored as a critical asset.

Data governance initiatives are impacted by the formal and informal

structures and cultural components that make up an organization. The formal side of an organization relates to structures, rules, regulations, processes, procedures that are mandated either in writing or through established norms. The formal side imposes and establishes how employees must engage with key stakeholders, both internal and external. The informal side, on the other hand, is where the real values, beliefs, priorities, expectations, and acceptable code of conduct is evidenced. It is obvious that there is no clear demarcation between the formal and informal side. Both are intricately interwoven, and employees may not even notice when the formal side overtakes the informal or vice versa. However, from a data governance perspective, this demarcation is important and consequential. The tenets of data governance demand that employees display consistent behavior at all times, whether it relates to the way an organization does business formally or informally.

Data governance at its core refers to the knowledge that an organization has about its data. It refers to the organizational framework, which includes policies, procedures, and processes, to create, analyze, leverage, share, protect, and regulate organizational data for the larger good of all stakeholders, including customers, employees, vendors, and partners.

Organizational culture is reflected in the way people perform tasks, set objectives, and administer the necessary resources to achieve objectives. Culture affects the way individuals make decisions, feel, and act in response to the opportunities and threats affecting the organization. (Wikipedia)

Hence, initiating or attempting to execute data governance initiatives without understanding the culture is equivalent to driving a car without knowing how to start the ignition.

Organizations are defined by cultures and subcultures. Cultures and subcultures are influenced by the external and internal environment, the industry in which the firm operates, the size, demographics, and skill level of its workforce, and the organization's responsiveness and strategic approaches to past and current challenges. These are just a few of the factors that influences organizational and unit culture. Just as how human beings have multiple personalities and just as how individuals wear “different hats” depending on the role they play, organizations also have multiple personalities. The larger the organization, the more likely it is to have many subcultures.

The organizational culture defines in broad overarching terms the values, beliefs, shared assumptions, behavioral norms, and other generally accepted practices. The subcultures describe and define variations within the above mentioned parameters and, in some cases, may even be entirely different or conflicting with the core elements of a given organization's culture. Hence, from a data governance perspective, it is essential to understand not only the organization's culture but also its subcultures.

Cultural Models

There are many lenses through which we can study how to successfully implement data governance within an organization's culture. There is no one best way to do this. Looking at different models to assess culture may provide a richer perspective of an organization's culture and its subcultures. We have selected a few models that are most appropriate and valuable from a data governance perspective.

Multidimensional Aspects of Culture

Culture is not unidimensional, but multidimensional. It has to be studied like a 3-D model with many moving parts. For example, Cameron and Quinn (2011) explore organizational culture along these dimensions:

Dimension One

1. **Internal maintenance:** Are all systems and processes functioning smoothly internally? Are things appropriately interconnected to avoid silos?
2. **External relationships:** What systems, processes, and approaches do we deploy to address the competition and to differentiate ourselves from the crowd?

Dimension Two

1. **Organic processes:** Are we flexible and dynamic in our processes and response time so that we can adapt to market changes and unexpected

turns? Are we agile and adaptable?

2. **Mechanistic processes:** Do we have processes in place to avoid chaos and confusion? Do our processes encourage stability and control and prevent unnecessary churn and meaningless work?

Hence, the following models should be viewed holistically from an internal and external perspective, and by understanding the organic and mechanistic processes by which an organization conducts business.

Mission, Adaptability, Involvement, and Values Cultural Model

The following four dimensions can be useful to understand organizational culture from a data governance perspective:

1. **Mission:** This is a critical foundational element to understand any organizational culture. What is the primary living mission of the organization? This is not about what it states on the website or glossy brochures. Mission is revealed in the key drivers of organizational performance and focus. Where is the organization heading? What are its strategic priorities? What do employees really get rewarded for? If employees don't know or understand the difference between strategic priorities and everyday tactics, they will focus on the easy part, which is tactics, rather than strategy, which is difficult. We know that strategy drives tactics and not the other way around. Hence, from a data governance perspective, if we want employees to focus on strategy, understanding the real mission, not the stated mission, becomes critical. Data governance initiatives should revolve around the *real* mission of the organization.
2. **Adaptability:** Employees must adapt to changing internal and external circumstances, which demand continuous learning. But the truth is employees don't always adapt or change or may even want to learn because of existing reward structures. Resistance to change and learning may also be driven by the subculture within an organization where some units are more prone to resisting while others are prone to learning,

creativity, and embracing change. Clearly it is ideal to launch data governance initiatives with units that are most open to change and adaptability.

3. **Involvement:** Select and work with units where teamwork and empowerment are high. When employees are involved and engaged with the internal and external change dynamics and are willing to meet head-on the dynamic nature of the workplace, data governance is likely to succeed.
4. **Values:** The greater the alignment between organizational values and personal values, the greater the propensity to follow the leader. There is no easy way to gauge this factor, but observing employee's behavior will shed some light on this important element.

Clan, Adhocracy, Market, and Hierarchy Cultural Model

1. **Clan:** In a clan culture, knowledge is shared extensively. There is great emphasis on standardization and the focus is on collective learning. There is a deep interest in mentoring employees so that everyone understands the importance of why things are done in a certain way. Work is regarded as an extended family. People generally follow the leader. The downside is that in clan cultures, “this is the way we do things here,” can be a roadblock. From a data governance perspective, change may be slow, systematic, and structured, with long hours spent explaining the policies and procedures and providing detailed justifications. However, since the clan culture values collaboration and teamwork, once processes are in place, the chances of long-term success are also high. Data governance leaders must prepare for pushback. The Japanese are cited as leading with the clan culture. Many multigenerational family-owned businesses may also embrace the clan culture.
2. **Adhocracy:** In an adhocracy, speed matters. Action is prized. Agile learning is emphasized. Employees are energized by change, speed, quick response, and action and hence rules, regulations, and processes

may be considered as burdensome and getting in the way of progress and success. A spirit of creativity and entrepreneurship in these organizations translates into strategic risk-taking, innovation, and a high desire to lead the pack.

3. **Market-oriented:** This has considerable overlap with adhocracy cultures. But unlike adhocracy that is more entrepreneurial in its focus, market-oriented follows the dictates of the market. It is highly tuned in to competitors' moves and is focused on the bottom line. Tools, techniques, systems, and structures that strengthen the ability of the organization to aggressively compete in the marketplace are valued.
4. **Hierarchy-oriented:** In this organizational culture, structure, control, and monitoring rule the day. Rules and regulations as well as policies and procedures are treated as tools to control employee behavior with the intent to increase efficiency and effectiveness and prevent loss, theft, fraud, and wastages. Bureaucracy is treated as an essential precursor to order and efficiency. The emphasis is on getting things right the first time. It ensures that decision-making is guided by a roadmap that all employees are encouraged to follow. This approach works when external markets are stable and predictable.

Power, Role, Task, and Person Cultural Model

1. **Power culture:** As the name implies, in such organizations power resides at the top. The tenets and extent of power are clearly communicated and are widely understood. Decisions are made by the power brokers, and any input from outside the power circle may or may not be considered. Hence, from a data governance perspective it is important to start top-down and ensure that the senior leadership is on board not only with the strategy but also with the overall tactics of strengthening data governance.
2. **Role culture:** In this culture, the title of the role and the hierarchy of the role within the organization define the decision-making power and authority. Structures are highly defined and carefully preserved in such environments. Power is based on position and role definition, and not necessarily expertise or skill level. Control and strict adherence to rules

and well-established processes and procedures are considered essential to do well in these settings. From a data governance perspective, it is essential to begin with assessing the relationship between power and roles and recommending structures that further strengthen existing control and monitoring paradigms.

3. **Task culture:** In this culture, expertise overrides role and titles within the organizational hierarchy. The goal is to bring the best minds together to solve a given problem. The greater the expertise, the more power the individual or the team is given to move the organization forward. Ideally suited to a matrix structure, such cultures get their work done through expertise-based small teams. Data governance, in such environments, succeeds if the team leading such efforts is perceived as trustworthy experts.
4. **Person culture:** In a person culture, the individual drives the culture and sets the tone for how problems are to be solved. Such individuals may have real or perceived power and it is up to the individual to determine how he or she wishes to use it and for what purpose. Law firms are often cited as an example of such cultures. From a data governance perspective, this can be a highly challenging environment. Data governance may have to be implemented on a piecemeal basis.

Our goal in this section was to introduce the reader to a few models that help to define the culture of an organization. The above list is not comprehensive. The idea is not to peg the culture of a given organization into one model or the other. It is designed to provoke a few insightful questions to help identify an organization's culture. This is essential for multiple reasons. First, the data governance initiatives that worked in one organization may not work in another precisely because of cultural differences. While data governance experts may be tempted to implement their knowledge and expertise about best practices in data governance in their current organization based on their experiences at another, best practices must be highly tempered and customized to current cultural norms, both visible and invisible, stated and abstract, to be successful. Second, culture is not uniform even within an organization. Every organization has subcultures within its departments or business units. Even teams have their own subcultures. Hence, it is important to observe the subculture of each group or team when it comes to data governance. Understanding the organization's culture will play a

huge role in how the mission, vision, goals and objectives, and governance processes and priorities are communicated. In some subcultures the preferred method may be, “Tell me what you want me to do and get out of my way,” while in other cultures employees may expect a say in how rules and processes are formulated. Third, the critical elements of data governance, including data collection, data integrity and quality, data usage, and data sharing protocols, also have underlying cultural components which must be understood. Finally, culture and subculture determine whether employees have a long-term view of the organization and their role as a data governance advocate or not. Some cultures encourage short-timers' attitude, while other cultures encourage employees to work for the larger good.

Organizational culture is the “soft side” of data governance. Hence, data governance advocates, under pressure to deliver, may be tempted *not* to invest time in understanding the organizational culture. However, the benefits of gaining deep insights about the culture and subculture that permeate an organization can yield rich dividends. This understanding comes not from theory or academic models, but from intense observation of how people behave, in general, with internal and external constituents. Based on this observation, we recommend tweaking or customizing two critical components: communications plus monitoring and control.

Finally, how we communicate with key stakeholders is strongly influenced by cultural norms. Communications is intimately tied to how people behave within a given cultural framework. We emphasize that communications must be customized to the cultural norms exhibited by a group or a unit. In the absence of that, an approach that was successful with one unit may be a terrible failure in another unit. The extent of monitoring and control needed to normalize proper data governance behaviors is also influenced by culture. In highly collaborative teams, excessive monitoring and control may lead individuals to question and eventually, work around control processes. In such cultures, employees may feel they are not trusted and push back. In dysfunctional cultures, monitoring and control may be highly critical for survival and success. We believe communications are so vital to data governance that we have dedicated an entire chapter to it (Chapter 3).

In short, we state two important messages:

1. Pay attention to culture and subcultures.
2. Be prepared to tweak any communication plans and approaches to get

the best out of every subculture.

Change Management

As we have stated before, data governance initiatives are not like projects with a finite start and end date. Data governance is a project with no end date and perpetual scope creep. It is an essential initiative that every organization must pursue, regardless of its size, industry, or market. This requires systematic and ongoing fixes, updates, enhancements, modifications, and deletions. It requires creating new policies and procedures as the regulatory environment changes. It demands that lessons learned, and there are plenty of lessons to learn, be incorporated into existing structures. In other words, data governance requires acceptance of the fact that it is like a moving vehicle, one where the landscape and the destination are ever-changing. The management of change entails taking a structured, systematic, collaborative, and measurable approach to achieve a set of predetermined goals.

The term “change management” is a popular and commonly used term in the business world. Yet, it means different things to different people. In its simplest form it refers to managing the inevitable change that all organizations undergo on an ongoing basis. Change can be small or big, planned or unplanned, organic or systemic, strategic or tactical, or operational or mission-related – regardless of the nature and scope of change, every organization undergoes change. Sometimes the change has a stealthy slowness that it may go unnoticed for a while. Regardless, all organizations are in the throes of change, no matter what the speed of change may be. And, organizations that reject change take a shortcut to loss and eventually, death.

Change management and data governance go hand in hand. Why?

1. Data are dynamic, and any corporate asset that is valuable and dynamic must be carefully managed.
2. People managing the underlying policies, procedures, and processes must be advocates and stewards of the changes that impact and influence data and decisions driven by that data.
3. Projects are often the platforms through which organizations achieve their business goals. Ensuring data integrity, quality, accessibility, usage,

and security across all projects and platforms is essential to avoid issues of inconsistency and duplication. This means that projects must be embedded with core best practices in change management when it comes to data and data governance. Without this underlying foundation, any project can undermine the overall data governance efforts and initiatives in an organization.

4. The policies, procedures, and processes underlying data governance itself need to be guided by a change management plan and strategy. In the absence of a well-planned change management philosophy, ad hoc and arbitrary policy changes can lead to unintended chaos.

These are a few of the compelling reasons why change management and data governance are interwoven. Data governance at its core is about continuous improvement efforts to change and enhance the value of data as an asset.

Although the reasons for embracing change are compelling, this does not mean that change is easy or that change initiatives are accepted and embraced by employees. Change is hard and people resist change for valid reasons and sometimes for reasons that are hard to understand or decipher. There are many competing and conflicting forces and agenda in any change management initiative. This is a given. To expect that everyone will buy into the idea that data are an invaluable organizational asset is setting oneself up for failure. It falls on the shoulders of the data governance team to understand the power of misaligned agendas and find strategic ways to navigate the landscape.

Effect of Culture

We present below a framework to help data governance change agents to understand and prioritize some of these competing forces.^{iv} The Competing Values Framework by Quinn and Rohrbaugh captures the essence of balancing competing forces: internal focus vs external focus, control, and flexibility.^v This is an extension of the model that we used earlier to discuss the four types of organizational cultures: clan, adhocracy, market, and hierarchy ([Fig. 1](#)).

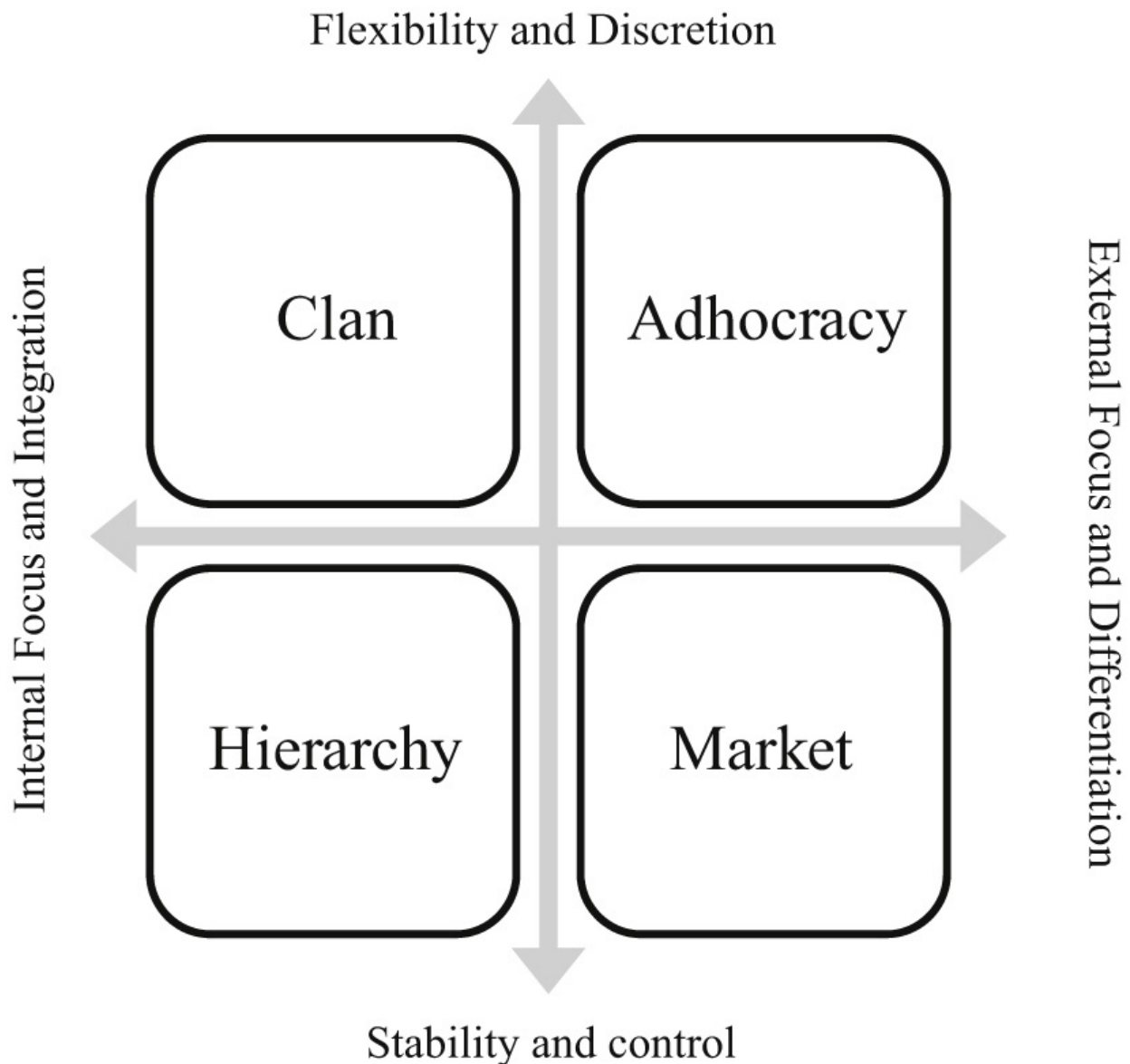


Fig. 1. Dimensions of Organizational Culture Based on Quinn and Rohrbaugh (1981).

The four types of organizational culture that we discussed using this model differ significantly along the two dimensions of organizational focus (internal vs external) and organizational structure (control vs flexibility). This struggle with competing values is not unique to a few organizations. Instead, it constitutes the essence of the nature of any organization. Change management initiatives must then embrace this set of conflicting values and focus in order to explore and identify a good balance. In the case of data governance, this becomes even more

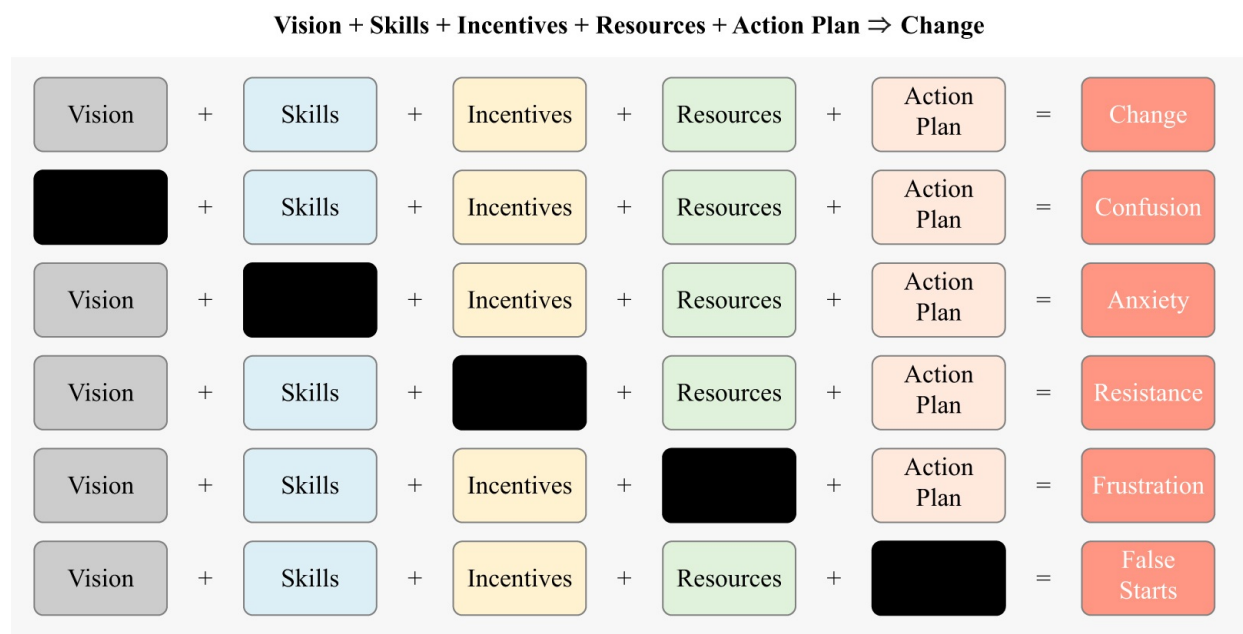
important. Flexibility and control must coexist. Data governance policies and procedures must take into account the internal world of the organization and its external stakeholders. Hence, it may be helpful for data governance leaders and advocates to explore each initiative within the background of the Competing Values Framework.

Data governance is inherently political, and the above frameworks summarize this fact nicely using a business framework. It demands compromise and negotiation and the willingness to be in the game for the long run while operating within the constraints of the organization's culture. Data governance must find its foothold within the formal structures established by the organization, such as policies, procedures, regulatory requirements, compliance mandates, internal processes, bureaucratic controls, and informal structures, such as behavioral norms, individual choices and judgment, beliefs, and values. While change can be formalized through formal structures, its success demands that informal structures work in sync with the former. Hence, a realistic approach is essential to data governance success – an approach that calls for deep sensitivity to competing and conflicting agendas and stakeholders.

Managing the Process

Once the cultural issues are acknowledged, understood, and planned for, an organization needs to actually undertake to make the change. Knoster, Villa, and Thousand (2000) outlined the components necessary for change to be successful, and these have been adapted hundreds of times and we will do so again here.^{vi}

There are five components that are essential for change to take place, and this framework is helpful for identifying what is missing if a data governance change isn't successful. We'll use some basic arithmetic representations for this section. [Fig. 2](#) shows the relationships.



Source: From Knoster, Villa, and Thousand (2000); adapted from Ambrose (1987) with permission.

Fig. 2. Components and Results for Managing Complex Change.

If any of these pieces are missing, change will not be successful. We'll address each component as it relates to data governance in turn:

Vision: Without vision we have: **Skills + Incentives + Resources + Action Plan ⇒ Confusion**

Understanding why a change is happening and what an organization hopes to achieve with that change helps to provide clarity for the outcome. Without understanding why a new data standard is being implemented or a concept is being defined or a data stewardship activity is being planned, employees will be at a loss to understand how it relates to their everyday work. Being able to articulate the vision for that change, not just for data governance as an abstract concept, will be important for having key stakeholders in alignment. A data standard promotes consistency across the organization and helps ensure the same data are being used to explain the same concept. A documented definition for a key concept helps to ensure that communication is clear and on point (see Chapter 3 for more examples). A data stewardship activity builds community and provides important training for a potential new activity. These simple explanations help employees see the important “why” for any proposed change.

Without that, the outcome is likely to be confusion.

Skills: Without skills we have: **Vision + Incentives + Resources + Action Plan** ⇒ **Anxiety**

Ensuring that the people who will be charged with doing something different have the right skills for the new task can help keep employees focused and engaged. If a new data visualization tool is being deployed or a new metric being created to measure a key output, employees will be much less nervous about the change if they feel they understand and can master any skill required to use the new platform or process. Not preparing affected staff for the change and ensuring they have the skill necessary to be successful after the change is implemented will lead to anxiety.

Incentives: Without incentives we have: **Vision + Skills + Resources + Action Plan** ⇒ **Resistance**

Providing insight into how an employee will benefit from a change is key to gaining his or her buy-in. The “what's in it for me?” question that is often bandied about should be adequately addressed to help employees feel vested in making the change successful. Explaining that the new data standard will make combining data from different departments easier and save them time may temper the pain of having to do the work to adopt the standard. If the new data visualization platform provides better output that is more suited to what they need to do, acknowledging that will help reduce the feeling of loss that comes with going from being an expert in the old platform to a novice on the new. Without linking change back to a benefit to the employee, no matter how slight or obscure the connection, the change is likely to be met with resistance.

Resources: Without resources we have: **Vision + Skills + Incentives + Action Plan** ⇒ **Frustration**

At some point, everyone has faced the impossible situation of being tasked with doing something and not having the time, talent, or technology with which to do it. Successful change requires some investment. The new standard might require reprogramming systems or changing processes, potentially requiring time and money. A data stewardship activity that may add additional requirements to a potentially full employee workload so there will need to be trade-offs or the newest request is likely to be the first thing to get overlooked. Without properly supporting a change with additional resources, staff and leadership are going to be frustrated.

Action plan: Without an action plan we have: **Vision + Skills + Incentives + Resources** ⇒ **False Starts**

Even with the best vision, skilled staff, aligned incentives, and an abundance of resources, no change can truly take hold without a plan to do so. An organization's culture may shape what a plan looks like from abstract to detailed but without some kind of roadmap, a traveler will never get far from home. It is key to outline how the new standard will be agreed upon and implemented or when the new data visualization tool will be available with details on how and when the transition will take place. Without an action plan to help guide even the most willing and able employees to the end goal, any attempt at change will materialize as a series of false starts.

Data-driven Culture

Datafication – transforming all things under the sun into a data format and thus quantifying them – is at the heart of the networked world. Just as electricity changed industrial processes and domestic practices in the nineteenth century, a data-driven paradigm will constitute the core of twenty-first-century processes and practices. (From The Datafied Society: Studying Culture through Data)

Our focus in this chapter was to show the importance of understanding organizational culture in order to achieve the goals of data governance. The success of data governance relies on creating a meaningful data culture, where data culture is defined broadly as the cultural elements that define how an organization values data. What are some of the outstanding characteristics of a data-driven culture?

1. The organization protects and treasures data as it does any of its other valuable assets. It treats data as a resource that is vital for its financial health and survival.
2. It has a strong understanding of the relationship between data integrity and quality and the integrity and quality of its decision-making. In other words, it views data integrity as fundamental and essential to making effective decisions.
3. An organization with a rich data culture expects employees at all levels, from entry-level employees to its C-suite, to treat data as an invaluable

asset and to take a data-driven approach to problem-solving and critical thinking.

4. Data and data-driven decision-making are treated as competitive tools that help the organization achieve a competitive edge in the marketplace. Hence, data drive the entire gamut of decisions, from strategic to tactical.
5. Employees are provided intense training in how to create, store, leverage, and protect data. In the absence of meaningful and ongoing training, employees may not fully understand or appreciate the importance of data to the organization.
6. In a data-driven culture, intuition and gut feeling do not replace data or the truths hidden in data. Instead, these two go hand in hand.
7. Top executives and senior managers in data-driven cultures understand the immense value of data and their power to transform companies and engage key stakeholders. They walk the walk. They invest in technology and training. They are continuous learners who keep abreast of the latest developments in data governance. They invest the resources necessary to make an impact. They communicate in strong and uncertain terms the importance of treating data as an asset. They understand that this is an ongoing initiative, not a project-based one. They understand failures are inevitable and stand by their team when things go wrong. They establish performance metrics to reward data-driven decision-making.
8. Data-driven organizations are strategic in how they think about data. Their approach is not ad hoc or arbitrary but is grounded in a pragmatic and progressive approach to the use of data to strengthen decision-making. Because of this approach, the chief data officer is often an integral part of strategic decisions.
9. In data-driven organizations, the entire management team is tasked with breaking down silos. This results in greater collaboration and openness to sharing data for the benefit of the entire organization and not just one unit or team.
10. In data-driven organizations, communication is frequent and consistent.

There are no surprises about policies or procedures or processes. Stakeholder input is valued and decision-making around data creation, usage, and sharing is collaborative.

11. Data-driven organizations avoid jargon and help people from nontechnical backgrounds understand the role and importance of data governance.

The bottom line is a data-driven culture that does not happen overnight or by accident. It is the result of a seeker's mentality to understand the real value of data to the growth and success of an organization. It is the result of building a culture that understands, appreciates, and is rewarded for data-driven decision-making. It happens because the organization has invested resources over a long period of time to ensure that it stays on the cutting edge when it comes to its ability to deploy its data to achieve a competitive edge. It is successful because the difficulty of change is acknowledged, the effect of culture on change is understood, and the steps to manage change are taken and clearly communicated. In the chapter on data maturity, we explain how data culture is strengthened as the organization moves up on the data maturity scale. Hence, no matter where an organization currently is in terms of its data culture, there is opportunity to grow. Data governance can play an instrumental role in helping organizations succeed in building a strong data culture.

Conclusion

Data governance does not happen in isolation. Instead, it is deeply embedded in the roots of an organization's culture. Data governance initiatives can only succeed if they are customized to an organization's culture. We know that culture is often not easy to describe. In this chapter, we presented a few models to assess an organization's culture. It is difficult, if not impossible, to fit an organization's culture neatly and squarely into any one model. This chapter is designed more as a reflection exercise to consider the different elements that constitute an organization's culture and see where one fits. It is an ongoing reflective exercise.

We also noted the importance of subcultures within a given organization. Data governance initiatives should take the subculture of business units and teams into account. Without this what worked in one unit may or may not work in another. Further, cultures and subcultures determine how best to communicate

data governance initiatives with key stakeholders and employees within a large sphere of influence.

Organizational cultures and subcultures form the foundation for creating a strong data culture. We listed several tenets of an organization that has a progressive data culture. While the list is not comprehensive, it provides some key characteristics that distinguish a data-driven culture from one that is not. Further, a data-driven culture does not happen overnight. It is a work of creativity, innovation, persistence, and strategic thinking. As an organization moves up the data maturity scale, its data culture also grows and strengthens.

Finally, data governance and organizational governance are complex. This requires a strong change management framework that considers not only the business and technical requirements but also vigorously incorporates a robust understanding of human behavior. Data governance impacts everyone in an organization, and hence it is essential to involve and engage individuals, teams, and leaders in change initiatives. Sustaining the capacity for change is not easy, especially in complex projects and rapidly changing business environments. Finally, data governance starts from the top. Senior leaders must be active advocates of change. They should build a culture that rewards data-driven decision-making and actively embrace change. Hence an in-depth understanding of change management is vital for data governance success.

Questions to Consider

1. How would you describe your organizational culture? Pick five keywords and explain how they capture the culture.
2. Which model most closely reflects your organizational culture? Why?
3. How does your organization handle change management? Has it been successful? Why or why not?
4. Describe what a data-driven culture would look like in your organization. List five specific things that would indicate that your organization's culture was data driven.

3

Communication: Key to Success

Introduction

As mentioned in the previous chapter, language is an important part of culture, including organizational culture. Books and journal articles abound discussing the linkage. Think about all the acronyms, buzzwords, and other business lingo that is prevalent at each institution, some of which may be common to other organizations and some of which may be found at no other. Most organizations have an internal language, or dialect, and common understanding of that language is necessary for the organization to function appropriately. This is true for data issues as well. To achieve a common data culture as outlined in the last chapter, everyone needs to be speaking the same language.

A common mistake is to assume that because different people are using the same words, they are using it to mean the same thing. That is often not the case. In addition, it may be that people are using different words that actually mean the same thing. The more decentralized an organization, the more likely that there is a lack of a common understanding of key items, concepts, or data issues. If an organization is new to data governance or has new data activities occurring, there may not be any language to effectively communicate these new things. This kind of language differential is going to be problematic to actually get a coherent and cohesive environment for data governance.

Foundational Work Is Necessary

Talking about what words mean often seems painful and rudimentary. Is it really necessary to have a group discussion about what everyone means by

“application?” Isn't it possible to just “do data governance” or “do analytics” without going back to kindergarten? If the human resources department is using “application” to mean the package of information that is collected when someone wants to be hired for a job and the information technology department uses “application” to refer to a technology platform, then there may be serious challenges in getting any agreement on how to collect data on “applications.” While this example seems extreme, it is meant to showcase how different perspectives provide different contexts for a word's meaning. For data governance to be successful, it's really important to make sure the people have the language that they need to be able to discuss things coherently. In some instances, it may seem as basic as ensuring that the concepts that the data capture are being described consistently by everyone who deals with those data.

In other instances, it is important to make sure that people have the right vocabulary to talk about the data themselves as opposed to the concepts that the data are capturing information about. In some organizations, staff may not know how to describe a collection of data objects. There may not be a common way of indicating a single data element. As data requirements and availabilities change, there may be a need for words that haven't been needed in the past or there may be some words from the past that should no longer be used. A common language requires that people know what words to use when and that they know everyone is using those words in the same way.

General language issues notwithstanding, here are a few truths about language that will be key to a successful data governance program.

Definitions Matter

The first step in creating a common language is knowing what words mean so that we can compare definitions. It is unwise to assume that two people are working with the same definitions. This holds for even the most basic objects or things. Let's take a very concrete example: defining how to capture information about buildings. Since this is such a basic concept, it will be tempting to assume that everyone knows what a building is.

It might be surprising to find that there isn't general agreement for something that seems so simple. Even dictionaries don't necessarily agree: One publication defines a building as “that which is built; a structure, edifice.”^{vii} A second defines a building as “a usually roofed and walled structure built for permanent

use (as for a dwelling).”^{viii} Yet a third definition is “a roofed structure for permanent or temporary shelter of persons, animals, plants, materials, or equipment.” All three are similar with differing degrees of detail. However, there will be things that are “structures” that do not have roofs and are not used for temporary shelter. They are buildings by the first definition yet not the third. It is important to decide what is relevant under which circumstances.

Organizations that have multiple locations may need information about where the employees, equipment, and other assets are located. Those who are trying to manage data about these locations need to make sure that the people who maintain the buildings and the people who deliver packages to the buildings have the same definition of what constitutes a building. For organizations in some industries, any of these three definitions might be suitable for capturing data on buildings. For institutions of higher education, however, only one of the three definitions given is relevant. The third definition is from the *Facilities Inventory Classification Manual* published by the National Center for Education Statistics and it's the one relevant to colleges and universities in the United States. If a college or university wants to count the items that are classified as a building, there is additional information and restrictions that need to be considered. In order for something to be classified as a building in an inventory, it needs to be attached to a foundation, roofed, service by utility exclusive of lighting, and requires significant maintenance and repair activities, a definition significantly different than the basic concept of the building.

Even for the most commonly used terms, well-documented and well-communicated definitions are critical for building understanding across the enterprise. While it may seem rudimentary, taking the time to craft definitions and share them across data silos is a foundational part to ensuring buy-in, understanding differences in business functions, and knowing when different words really need to mean different things.

Context Matters

While definitions are critical, they may not be unique. Dictionaries often list multiple definitions for a word because context matters. The “application” example used previously illustrates this case well: different parts of the organizational chart have a different context for the term “application.” A common context in general language may be geographic in the form of regional

dialects or regional words for things: There are different areas that describe things with different words as part of the local dialect. That is not to say that people don't understand the common words for those things, they just use different words depending on their geographic location.

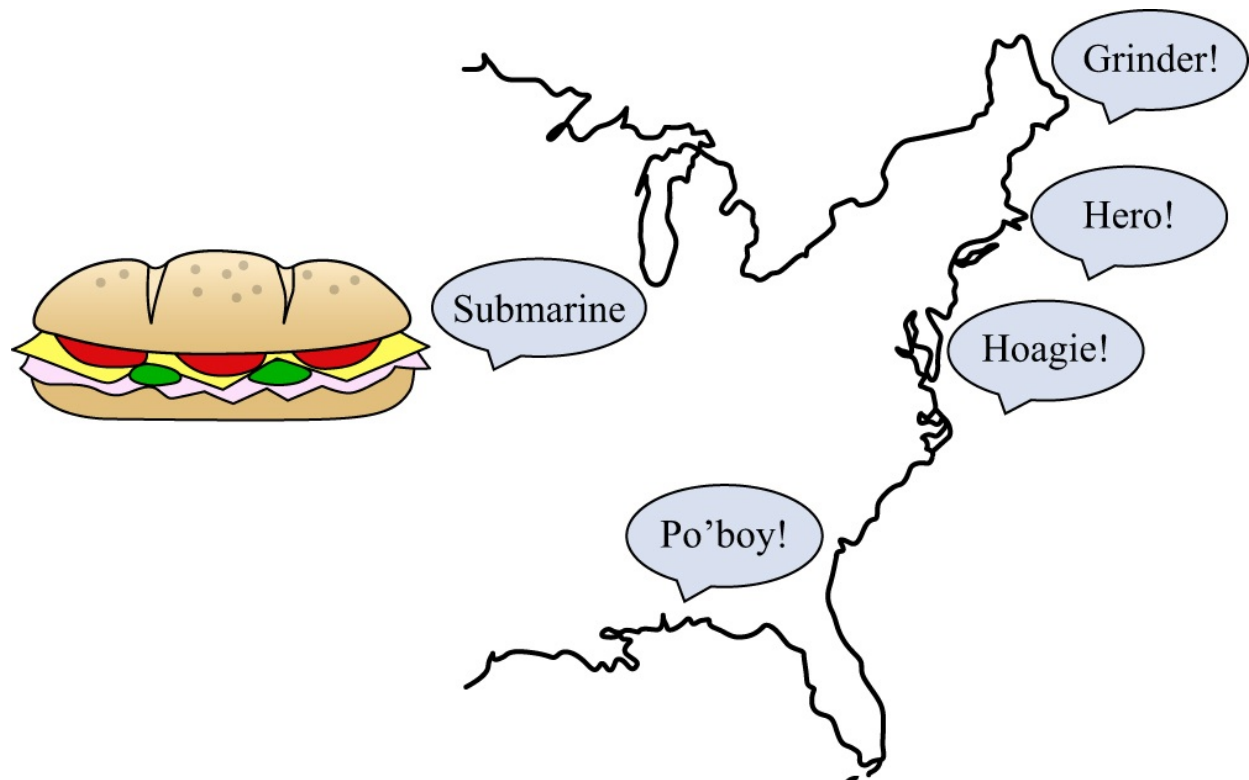


Fig. 3. Regional Differences in Terminology

Consider the humble submarine sandwich. In New England that would be called a grinder. In New York, it might be called a hero. Around Philadelphia it's called a hoagie. And in the south, there's a sandwich that is very similar that is called Po'boy ([Fig. 3](#)). There's lots of different names for the same thing, if they can be considered the same thing. According to the dictionary, a Po'boy and a hero are synonyms: they are both listed as part of the definition of submarine sandwich. People from the south are likely to have a very different opinion.^{ix} For each area, the context is important.

While sandwich monikers are not relevant for data governance, the regional differences in terminology are analogous to siloed differences across an organization. This dichotomy can be especially pronounced when dealing with technology implementations. In places where the definitions might be consistent,

the name of the data element for that concept in one system might be completely different than the name in another system. If everyone agrees on the definition of “customer” and uses that definition appropriately, the marketing platform might store that information in CUST_NAME and the finance system might use the word PAYEE. Even if everyone agrees they are the same thing, if two units don't have the same name, any references to these terms between the two departments can be very difficult.

Standards Matter

Dialects and regional differences are important and are not going away. So how are we meant to develop a common language? First, there must be agreement on a standard definition. Standards are misunderstood, as is data governance in general, as the stick that forces people to conform. Instead, they can be viewed as the common reference point. If everyone understands what a submarine sandwich is, regardless of their regional term, then discussions about that concept which use that term can avoid a great deal of confusion that is created by mixing po'boys and heroes.

The stick analogy would be relevant if there is an expectation that the standard representation would be forced on local populations that might use a different word. The organizational culture, such as the power or hierarchical culture, may dictate such a move. For other organizational cultures, acceptance of a standard may be much smoother if there is an agreement that the standard vernacular is used across regions (or across departments or across organizations) to assist and enable conversations without dictating changes to the local dialect. It may be far more effective to build a common data culture by not insisting that local communities use the standard term. Understanding the dialects and understanding how they are related is crucial for building an inclusive data governance framework and in getting broad adoption across different parts of the organization. Let the people eat their po'boys and their heroes at home but make sure everyone knows what they get when they order a submarine sandwich. Nothing kills an idea faster than forcing it on people who don't see the value.

This is especially true for technology implementations. Using our previous example, the concept of “customer” could be called whatever it needs to be called in the variety of local systems in which it might appear, but when local systems share information or feed information to a data warehouse or dissemination hub, that's when the translation to the standard terminology or

standard field name should be enforced. This approach would imply that a standard representation of the concept in a data warehouse might be, say, CUSTOMER. And the marketing platform can map their CUST_NAME to that standard when reaching across systems, as the finance system could do with their PAYEE item. The need for a common label can be met by using CUSTOMER in cross-departmental activities without requiring the pain of reprogramming all the downstream applications that are based on CUST_NAME and PAYEE.

The cost of reprogramming local systems to remove dialects and enforce standard representations for concepts or field names is often one of the reasons that data governance efforts fail. There doesn't seem to be a good way to justify that kind of reprogramming and redesign cost and that decision gets translated into abandoning the concept of a standard definition altogether. The two should not be conflated. While there is an additional cost of maintaining a mapping, or crosswalk, it is generally far below either the reengineering costs of enforcing the standards in local systems or the cost of a failed data governance program.

Agreement Matters

That said, defining and agreeing on a standard definition is not necessarily an easy process. It can often require painful discussions and perceptions of loss of control. For someone in the organization, a standard will mean and change and change is hard. One way to help ensure agreement and adoption is to allow for the existence of other approved definitions defined relative to the standard. These versions can be thought of as **documented variants** and can help ensure clarity in differences. Let's say that there is general agreement that the definition of customer for an organization is “a person or organization that buys goods or services from our company.” Regardless of the local term, the definition applies across many parts of the organization, but not all. Perhaps the procurement folks have a valid business reason for defining customer as “a person or organization that we are negotiating with to buy goods or services from our company.” Small but important difference: if negotiations are lengthy or fail, there will be a period of time where the person or organization does not meet the standard definition of “customer.” This discrepancy might be evidenced in different reports or analytics about the size of an organization's customer base.

One solution that can save the standard definition is to allow a documented variant of customer that fits the procurement need. Note that without data

governance and the creation of a common lingo, the variant would exist, but it would be undocumented. By allowing the business need to be addressed and the difference to be clearly described, it avoids the confusion that would persist without the documentation and the discord that would persist without the standard definition.

Once a standard is finalized, the most important thing is to make sure that people understand what that standard is. This gives people a reference point to which they can compare their local term for the same object. Even if standards are not used regularly within local communities or system implementations, it is critical to have them. This helps to relate local terms to the standard term. Think of standards like the North Star: people don't actually use it to navigate because there are better options for that, but it can be used as an anchor to make sure that everyone knows where north is.

In order for this to work there needs to be widespread understanding and acknowledgment of the standard. If people don't have a word or a term for the concept that the standard is describing, like a submarine sandwich, they should be encouraged to use the standard instead of developing their own local term. However, if local terms already exist, crosswalks and mappings and other ways of linking the local terminology to the standard term could help bridge the gap and increase understanding across different departments or part of an organization with minimum confusion, cost, and resistance.

Documentation Matters

The death of understanding begins with a verbal agreement. How many meetings have ended with everyone in the room “agreeing” on a strategy or plan of action only to have the first email after the meeting describe something completely different than what some people thought they agreed upon?

Working on definitions and standards is no different. In fact, as we will discuss shortly, having written information can often help identify potential issues and diffuse tensions. Most organizations rely on a business glossary of some sort to help with this piece. The tool used to capture the documentation isn't nearly as important as the fact that it is captured in written form.^x No matter what the process is for determining the standard, or documented variant, sharing that information in writing is important for providing clarity and structure for the term. And if there is more than one term or definition necessary, document that

as well. As previously mentioned, there are very few words in an American dictionary that only have one definition. And there are just as few concepts that only have one term to capture them. Documenting the specific way that an organization uses a particular term or defines a particular concept is critical to creating a common understanding.

Politics Matter

Let's call out the elephant that is in most rooms: politics. All organizations face competing priorities and differing ideas which are often in conflict. How to define and use language is another area where there can be strongly held opinions and beliefs. While collaboration and consensus on standards for terms and definitions is likely to yield the best outcome, there are going to be times when an influential person or persons are going to decree that a particular word shall be used or particular definition shall hold. Business press often identifies this as the HIPPO management principle: the Highest Paid Person's Opinion holds the most power. Unfortunately, for many organizations such a situation is real and prevalent and needs to be acknowledged.

When faced with a situation where a standard is declared unilaterally, it may not be in the best interest of the data governance effort to fight that directly: acknowledge it, document it, and move on. The continued development of a transparent and well-understood common vocabulary can often help to minimize the effects of these politically motivated situations. The results may not be perfect but when people appreciate the importance of documenting and sharing information, it may become easier to highlight the difference in the way some decisions are made. At best, the transparency helps bring about the necessary cultural change. At worst, there is a record of how a potentially suboptimal solution came to be.

Emotions Matter

Because words matter, people are often very attached to them, or are repulsed by them. Whether it seems important or not, words often have emotional meaning to people. Don't assume that just because the conversation is about data and numbers the discussion is immune from emotional reactions. Suggesting alternative expressions or definitions can bring out seemingly irrational

responses. For example, there may be some people for whom the suggestion that their local term isn't the best one to be considered as the standard is something to be taken personally. Or it may be that the words that need defined or standardized for data collection are emotionally or politically charged in the current societal discourse, such as gender and race.

Regardless of the root of the emotions, it is important to acknowledge that people hold them. It is best not to ignore or try to minimize the reaction, regardless of how “silly” or “inappropriate” one might feel it to be. Remember, for people to speak a common language, there needs to be understanding for the feelings about that language. Avoid using phrases that involve the word “just” – as in “it's just a definition” – or “should” – “this should be easy.” Whether intentional or not, those words are loaded with a judgment that may not be shared by the person with the strongly held feelings. Be respectful of people's feelings while working to disentangle the reason for the emotion from the need to do the work that needs to be done.

When the Language Is Incomplete

The previous section focused on how to get agreement and documentation for the words and definitions needed to describe things about which organizations collect data. But what about the words to describe the data themselves? Or the people charged with data activities? For many organizations, this language is lacking, and leadership or staff may not even realize it is missing. It is difficult to know what people understand about data and what they don't without the words to express the important ideas.

Words to Describe Data

A big stumbling block for many organizations trying to operationalize data activities is lack of understanding or agreement on what data actually are or how to describe them. As mentioned in Chapter 1, one of the first challenges that any data governance program may face is the scope question: What are data? For many people, data are numeric representations of business concepts that can be organized into tables: data are what is found in a spreadsheet. For others, data are nearly anything born digital: text, audio, video, and numbers are all different types of data. If there isn't clear understanding and agreement about what data

are, then trying to govern those data will be an impossible task.

Some organizations might find that they have a de facto definition of data that is set by a predominant business line or area where data activities prevail. For example, some organizations may have data operations or data governance embedded or tightly entwined with their information technology operations. In this case, it is likely that much of the language used to discuss data is borrowed from the technology end. If data governance is a by-product of technology governance or implementation, conversations about data will naturally be conducted using words and terms related to the platform or application that houses or manages the data. If business lines describe their data pain points by describing problems with software that they use to access the data, it shows that they don't separate the data from the technology. As will be covered in more detail in Chapter 5, data are separate from technology and there needs to be a common language describing data regardless of the technology that is used to store them.

A first step might be to look at the highest level and decide how to name and describe **data domains**. Data domains are classifications used to exhaustively describe data. There are a few different ways to do this. The most common approach is to use subject or topical area. Subjects such as “customer,” “product,” or “finance” are often natural groupings to tie data together. These domains make it simpler to discuss issues relating to “finance data,” for example, which may make the topic more specific and the conversation clearer. Sometimes business functions are used to describe domains resulting in “marketing” and “sales” as groupings to describe data. Such data domains often result in duplication of data elements across domains. In such cases, written documentation of the domain definitions will be especially beneficial.

Once organizations decide how to describe the entirety of the data and the appropriate subsets, it will be helpful to carefully select words to describe specific assemblages of data elements. If domain definitions are based on subject areas, data elements that support a specific business process are likely to require a compilation of data elements from different domains. The purchasing function is likely to require data from the customer domain as well as the finance domain, for example. It may be important to use different terms to describe these different types of groups. Many people just use “data” as their collective noun (e.g., “customer data,” “marketing data”) but that is often too high level. For example, how do people know the right way to request access to data? Without a noun, they cannot make a well-formed request. Nor can the decision-maker

adequately describe what they are willing to provide. Knowing that someone wants “customer data” may be generically helpful but that doesn't indicate the scope of information being requested from that domain. In addition, it isn't clear how to reference the data grouping that has been compiled. Perhaps if the grouping is all from the “customer” domain, then people can easily discuss the “subset of customer data” that has been assembled. However, if the data accumulation has items from the “customer data” as well as the “finance data,” the subset moniker will not suffice.

Discussing collective nouns for data elements is generally omitted from data governance materials. Many of the standard reference or data governance books use “dataset” or “data set” (and sometimes both) as descriptors for groups of data elements but don't define either, assuming that the reader has the same definition as the authors. Like with the concepts that data are collected to quantify, clear terms and descriptions for data items are critical for enabling and encouraging a common understanding. Decide what terms to use, define them, and evangelize their use.^{xi}

Words about People

Organizations new to data governance may find that they also need words to describe the roles associated with data activities. As we will discuss in Chapter 5, any data governance framework has different roles and responsibilities associated with data collection, management, and use. Often these roles are associated with the various levels of data groupings (e.g., domains, subdomains, data sets) that are identified. The choice of label for these roles can affect the way people perceive the duties and reach. The following are some of the common labels for data roles.

- **Owner:** This descriptor has been in the spotlight as of late since the concept of ownership implies rights that an organization may not wish to bestow on an individual or department. It is becoming more and more common to hear the clearly stated principle that the organization owns the data. This helps prevent any person or unit from feeling or claiming that they have ultimate decision rights which can prevent data from being treated as an enterprise asset. If “owner” is a label that an organization chooses to use for a data role, the definition needs to be clear and the expectations for that role should be clearly described and documented.

- **Steward:** This term is generally used to describe a key role in any data governance framework. Because the commonly understood definition involves the safekeeping or management of someone else's property or affairs, the label fits well for data roles where ultimate decision rights are not vested. Steward is probably the most common label for a data role and is often applied at a variety of levels.
- **Trustee:** This label is fairly new to the data governance world so it doesn't have some of the historical baggage that terms like “owner” have accumulated. The meaning is similar to that of “steward” in that it involves taking care of someone else's stuff, but an important distinction is that a trustee generally implies being part of a larger group of others with similar responsibilities and so may be more appropriate for data roles that involve group decision-making.
- **Manager:** This label is often used to cover activities that tend to be more direct and operational than roles with the other labels. Often this term describes a role that is more actively involved with the data and with fewer decision rights.
- **User:** Application of this label is often overlooked but as data become ubiquitous across many organizations and the need for data literacy recognized, it is becoming more important to decide who the data users are and how that group is defined. Some organizations may decide that all knowledge workers are data users; for others, the term may involve only those who have specific analytics or reporting responsibilities. Deciding on who qualifies as a data user and what that means will be key to defining the roles and responsibilities that come with it.

Using the Common Language

Once the terms have been defined, standards identified, and data words deployed, it is important to decide how to best use the language to communicate with people about data. Nearly every data governance book, management text, and other publications give sage advice on what to communicate to whom, when, and why. We would be remiss not to provide some similar guidance. Instead of focusing on audience and strategy, however, we focus on the communication methods. As is discussed by any treatise on communication,

different kinds of messages serve different purposes and will find varying levels of success with different audiences.

Push versus Pull

It is very important to match the delivery mechanism for the information with the style of the intended recipient or no amount of slick graphics, carefully chosen words, and strategic stylings will convey the message effectively because they won't see it. In general, there are two ways that people consume information: it is “pushed” into their awareness (think email or text) or they “pull” it from a particular source (think websites or social media). Both methods can be effectively used to communicate about data issues, and rarely will one person be a user of both methods for the same kind of information.

Push delivery has two important qualities: immediacy and direct connection. When a message is delivered directly to the recipient, the content and the timing for the delivery will be tightly targeted. The author can decide on the specific audience and can share different versions of the message with different groups at different times. The drawback to push delivery is that it is essential to be able to identify all the individuals who need that information. People who are not on the email list but might benefit from knowing the content may not see it in a timely manner or may not see it at all. For some information, this limitation may be a benefit: maybe some information isn't appropriate for an audience beyond those involved in the immediate delivery. However, understanding how to reach the relevant stakeholders in general, and those with “push” preferences in particular, can be a challenge.

Pull delivery, in contrast, provides longevity and broader reach. Posting information on a website or other dissemination platform absolves the author from deciding who needs to know because everyone can have access, subject to security settings. This provides some consumer choice: people can choose whether to consume the information and when to do so. The downside to pull delivery is that it loses the direct connection with the intended recipient; presentation of the content needs to be more generic, and there is no control over when or if people get the information. Consider social media platforms. Consumers may or may not choose to watch the funny cat video that was posted. They may not even know it was there if they don't visit the platform in a timely manner.

For critical communications, consider using a push delivery even for pull

consumers. A compromise is to use a combination approach and push links to widely available content. Many collaboration tools and social media platforms provide alerts or allow users to “watch” certain kinds of content to be notified when there is new information. This setting provides push immediacy for pull content. However, it requires that people opt-in so this approach may not provide the breadth of notification necessary for some content. Authors may want to consider providing such services directly by using push mechanisms to alert relevant recipients to the availability of new information to be pulled for consumption.

Oral versus Written

Different audiences will require different delivery styles. While there is universal ridicule about the pointlessness of meetings, sometimes delivering the message in person can accomplish more than pages of written words. Especially for some “push” consumers, the message may not be relevant or urgent until a person is standing in front of them. On the other end of the spectrum, some “pull” consumers would do anything to avoid having to sit through “death by PowerPoint” and would much rather spend the time reading some content and having time to digest. For the time-pressed “push” people, a brief targeted message might suffice.

When determining the delivery style for specific content, consider the engagement of the people that are the target of the communication. Is it more efficient or effective to provide an opportunity to ask questions in person when receiving the message? Or is the audience likely to fixate on a small detail and derail the whole discussion? Will a group setting for a presentation enable constructive conversations among the participants? Or is a one-on-one conversation more appropriate for some stakeholders? Does the brevity of a slide deck provide the right level of detail? Or is the audience one that requires a great deal of background, data, or additional information that is best delivered in a document form? Like the push versus pull distinction, understanding the differences in the audiences will help to decide how to tailor the content to be delivered in the way most likely to be successful.

Inform versus Persuade

We do not mean to give the impression that how the information is delivered is more important than what the information is. We do mean to help provide practical guidance on how to make sure the messages are effectively received. In addition to the mechanism and style, the tone of the information helps for the message to be heard. To that end, consider whether the communication is meant to inform or persuade the recipient. Regardless of the style and content of the message, the language and expectations might be different. Providing information is an important part of looking to persuade, but it stops short of the goal of persuasion: getting people to do something. People often assume that informational messages are dry and factual and that persuasive messages evoke emotion; sometimes this is true but informational messages can also induce feelings and reactions. For example, the statement “The house is on fire.” is informational and rather staid. It could instead be written as “The house is on fire! There are newborn puppies in that house!” Now there is more information and more emotion, both stated through the exclamation marks, and implied by the imagined plight of the new puppies. Although emotional, it's still not persuasive because there is no implied action. The information is clear but there is no guidance as to what the recipient should do with the information about the fire and the puppies. Call the homeowner? Call the fire department? Rescue the puppies? If there is a specific action in mind, it needs to be part of the message so that there is a clear connection between the information and the intended action.

The previous example is hyperbole but serves to illustrate an important point when communicating about data governance: a broad group of stakeholders will be unlikely to be able to draw the same conclusion or infer the same immediacy without specific guidance. Unlike a house fire that is widely recognized as a disaster, the kind of challenges that creating and implementing a data governance program will face may not be recognized as critical by everyone in the organization without some guidance. In fact, in early stages of a data governance program, the issue itself might not be well understood so that more information may be required for people to even grasp that something is good or bad.

It may be necessary to provide information several times to ensure a good understanding of the basic issues before adding the additional persuasive language. For some organizations, the learning curve for data governance topics and issues can be quite steep. Informative messages may help move stakeholders far enough along that curve for them to be convinced that the action suggested is sensible and warranted. Such a strategy requires multiple communications on the

same topic but that will be necessary and should be expected.

Repeat versus Reiterate

One thing that cannot be overstressed: there is no such thing as overcommunicating about data governance. Even when great care has been taken to craft the message and deliver it in the most effective way possible, it will need to be repeated again and again and again. Sometimes it will be the same message to the same audience; other times changes to either the content or the mechanism will be necessary. Regardless, repeated communications should be expected. Like the language issues mentioned previously, not everyone is going to see and appreciate the issues like those charged with data governance. It might be the case that there needs to be repeated informational messages needed to get to a baseline level of understanding and appreciation before the persuasion will be effective. It might be the case that a great deal of information needs to be made available to be pulled before a push message is successful. Given that data governance touches all aspects of an organization at all levels, it will likely be necessary to use different delivery styles in different ways for different people and groups. It may be more effective to create versatile content that can be packaged and delivered in a variety of ways from the start, providing the right information to deliver at the right time in the right way to the right audience. And then it'll need to be done again. And again. And again.

Summary

Like many areas in the business world, communication is critical for effective data governance. As we saw in Chapter 2, data governance involves change whether the program be nascent or mature. Effective communication is crucial for change to be successful. We have posited that there are several key items that are necessary for communication to be effective. First, organizations need a common language to discuss the kinds of concepts that the data are describing. Building that common language requires many considerations around definitions, standards, documentation, politics, and emotions which we have described here. In addition, new language to help describe data roles and activities themselves may be needed. The data governance program may need to include teaching people words about data and about the people who will interact

with data. Finally, we presented some practical tips on how to make sure the language being used is effectively received by different kinds of stakeholders who have preferences for delivery mechanism, style, tone, and reiteration. All of these things can help ensure that communication is key to success.

Questions to Consider

1. Does your organization have words to talk about data? Is there an agreed-upon definition of “data”?
2. Is there a commonly understood language around data concepts? If so, how are they documented or communicated? If not, how might they be implemented?
3. What language does your organization use to describe data elements, collations, and roles? Is there clear documentation of these definitions?
4. What common types and styles of communication have been most effective when communicating with stakeholders in the past? What might you do differently?

4

Data Strategy

Introduction

By its very nature, developing an organizational strategy is intricate work because strategy is dynamic and interlaced with uncertainty and risk. Strategy demands the ability to integrate a diverse set of elements, functions, and even competing agendas into a comprehensive guiding document that is both functional and inspirational. The building blocks of data strategy involve more than just people, process, and technology. It involves the ability to continuously scan the environment and differentiate the important from the irrelevant. It demands the ability to paint an accurate landscape of a set of moving and often interdependent forces, while identifying potential landmines and volcanoes.

Strategy is both an art and science, the success of which is founded in human creativity, intuition, insights, and behavior, which is unpredictable and impulsive, to say the least. Strategy is a time-consuming and labor-intensive process that demands an intimate knowledge of one's organization and its people, the industry in which one operates, current market conditions, and local, national, and global forces. Strategy demands dual, yet competing, abilities, namely, to have one's feet firmly planted in current opportunities and challenges, while simultaneously soaring high to see what the future holds. The best strategists know that they have blind spots despite in-depth expertise in all of the above and therefore promote and encourage diverse thinking and relentless candor of all stakeholders.

The above challenges are no different whether it is strategy for an entire organization or strategy for the data that reside in an organization. Given the intricate and complex nature of strategy building and execution when it comes to organizational data, data governance leaders may be tempted to jump right into

execution mode while sidelining and shelving strategy for a later date. “I don't have the time to develop a strategy. I have to put out burning fires first!” is a common and understandable plea of leaders under pressure. This would be a mistake and a grave one, in any industry and for any organization.

Surveys and interviews with leaders reveal why the responsibility of leading data governance initiatives is so difficult. It is not because leaders don't understand its importance – at least most leaders do understand the value and importance of data governance and the risk of ignoring this field of activity. The challenge is to juggle the day-to-day pressures while keeping one's eye on the long term. The struggle is to make sure that at the end of the day strategy, organizational structure, operations and processes, people, technology, market forces, and regulatory and compliance requirements are perfectly aligned to help the organization achieve its mission and strategic goals. Data governance leaders who understand this complexity and can work efficiently and successfully within this framework will become indispensable to their organization.

Our goal in this chapter is to help data governance managers and leaders develop a meaningful data strategy by asking the right strategic questions. The purpose of this chapter is to train leaders to ask strategic questions, plenty of them, all the time! In many settings, the importance of asking the right questions has lower priority over finding answers to easy questions. Organizations that are focused and determined on finding answers to tough strategic questions will perform far better than those that take the easy way out. We offer pointers and guidelines to finding meaningful answers, but we don't offer detailed answers to any question because there is no one right answer. The answers and right approaches will vary from organization to organization, depending on its culture, data needs, data maturity, resident talent, resource availability, and political commitment to data governance.

There are many books and white papers that offer general guidelines and to-do lists on data strategy. The practitioner, however, needs tools and tips to develop flexible short-term and long-term strategies that continually evolve as the nature of the business and its priorities change. Grand ideas sound great on paper, but leaders in the field struggle to find ways and means to move the organization forward with demonstrated results, while being strategic in their thinking and actions, while articulating their vision and strategy to a diverse group of stakeholders.

What Is Data Strategy?

At any given point, on an everyday basis, most employees at different levels in an organization create, access, transform, analyze, use, store, share, and/or secure data. Data are the foundation for analysis and decisions, small or large. Data are the underbelly of countless decisions that employees are called upon to make every day. Data and the decisions they trigger are the currency for success in achieving goals and objectives and ensuring the sustainability and success of an organization. Data weave the past, the present, and the future into a cohesive tapestry to provide organizations with a powerful set of tools that form the foundation of competitive advantage. For the above reasons, data are both an individual responsibility and a corporate responsibility. Any individual who has any relationship with any organizational data has a data governance responsibility. However, employees cannot fulfill their responsibilities unless they have a clear understanding of the roadmap or the data strategy and the reasons driving that strategy. Employees must know unequivocally their data roles and responsibilities. They must have the tools, skills, and processes to fulfill their roles and responsibilities, and clearly understand the consequences of any negligent or callous behavior.

Data strategy is not about strengthening existing query and reporting capabilities and fixing a few broken links here and there in an arbitrary fashion. It is about methodically increasing the focus of the data use toward sophisticated analytics, ranging from predictive to unstructured data analysis, such as text and multimedia, data mining, utilizing sophisticated business and analytical modeling tools, building insights and intelligence through storytelling and dashboards, and natural language queries. This can only be done if the organization to begin with knows what data it has, where the data came from, where they are stored, who uses them for what purpose, the quality and integrity of the data, and how they are shared and secured.

Data strategy entails understanding the key elements and influencers that define and determine the success of any strategy. This includes (1) structural and cultural influencers – does the organization's structure support this strategy; (2) talent and resource influencers – are the relationships in place to execute collaboration and resource sharing; and (3) knowledge about processes, technology, and industry shifts – this is essential to execute any strategy. Finally, even if all the above mentioned structures and elements are in place, it is essential to understand and accept that execution of any strategy takes time.

Overnight return on investments (ROIs) are impossible and as the organization inches forward in becoming data-driven, patience and ongoing support of a dynamic data strategy is essential to long-term success and data maturity.

An analogy might be helpful here. Imagine creating a large quilt that is continuously and intricately being pieced together by quilters in different locations, many of whom do not always communicate with each other. The quilters may even have opposing interest and agendas, guided by a vision that is independent of the master quilter and designer. The skill level of each quilter may also vary significantly as each artisan displays different skills, talents, and models. Every quilter has their own pieces of fabric that they use to create their part of the quilt to capture a specific pattern. Not all quilters know the specific pattern of the other quilters, nor do they care.

The master quilter and designer is responsible for ensuring that the individual pieces being brought together from each individual quilter represent fine work with high-quality materials. There must be an accurate record of the source of each element that goes into the quilt. The final quilt needs to avoid duplication – an appealing pattern has cell constructed repetition, not haphazard duplication. The master quilter must ensure that this quilt meets its intended purpose and that all quilters execute their roles and responsibilities efficiently and effectively. Each quilter needs to communicate in a timely and consistent fashion with other quilters working on the fabric and they must share their knowledge, expertise, and lessons learned. The master quilter must ensure that as the fabric increases in size and complexity, it is protected from harm by internal and external elements. The job of a data governance leader is like that of the master quilter, only 10-fold more difficult and challenging.

Strategy and its many subelements and issues can be understood at an intellectual level. But in the ring, creating a data strategy that integrates and addresses diverse elements into a cohesive whole is an enormous challenge. We recognize these challenges and offer guidelines to start the process of building and updating an effective data strategy. There is no one answer that works for everyone. If world-renowned chefs were given the same set of ingredients, each one is more than likely to come up with dishes that are worlds apart. The same is true for data strategy – there is no one right approach or perfect answer. People, processes, technology, organizational culture, external and internal risks, compliance requirements, laws of the land, current data quality and maturity, volume and velocity of data, senior leadership support and commitment, and a whole host of other factors become input to building a foundation for data

strategy.

Take, for example, the US Army.

The scope of the data problem is striking: the Army intelligence enterprise needs to collect and exploit classified military intelligence; acquire and exploit publicly available information; exploit data collected by an increasingly large and varied array of sensors; and exploit data provided by allied, coalition, and international partners. As the quantity of data increases, so does the complexity in delivering a common intelligence picture to support collection, targeting, and decision making. Finally, the Army intelligence enterprise's transformation must comply with U.S. law and the various U.S. intelligence community, Department of Defense (DoD), and Army policies and regulations.^{xii}

The data strategy for the US Army will be distinctly different from that of a bank, or a hospital, or a global women's clothing retail chain. Therefore, the key to building a great strategy is to embed it within the business strategy of the organization and focus on using data to advance the strategic goals of the organization.

IT-driven Data Strategy vs Business-driven Data Strategy

It is easy to confuse a business-driven data strategy with a technology-driven data strategy. These are not the same. In fact, we find that many organizations treat data strategy as an IT strategy, often leading to implementation failure, loss of trust of senior management, and inability to deliver meaningful results.

A technology-driven data strategy focuses on how to leverage technology to achieve data quality, integrity, accessibility, and security. This approach leads to the data strategy being wrapped into the IT strategy for the organization. It advocates the view that data are an IT issue, not a business issue, leading to limited access to resources and a less than firm commitment from senior leaders. IT-driven strategy assumes incorrectly that the technical experts who manage data and all their related components have a strong and clear understanding of

how data are used to advance the business goals. This is equivalent to assuming that business users who use the data are technical experts who hold mastery over the technological aspects of data storage, integrity, quality, and accessibility. It is a wrong assumption.

A business-driven data strategy views data as core and precious organizational assets that are fundamental to the survival and success of a business. The elements of a business-driven data strategy are based on the data needs of the business and its ability to leverage data as a resource to increase profits, reduce costs, launch new products and services, innovate, thrive, and achieve market dominance. The language of an IT-driven data strategy is technical, while the language of a business-driven data strategy is the common language of the business – one that everyone understands. Hence, data strategy should always be a key component of the business-driven strategy. When the overarching business strategy drives the data strategy of an organization it sends an important message that data governance is a business function, not an IT function. When the data strategy is embedded in the business strategy, data assume their rightful place in the organization as a core and critical asset. For these reasons, data strategy should not be hidden or wrapped under the IT strategy for the organization.

This approach ensures that there is a common, universal language in the organization around data. Abstract, technical terms don't get in the way of buy-in or stakeholder engagement when business language guides data governance. In short, aligning the data strategy with the business strategy is the foundation for a meaningful and sustainable data governance program. It is important to ensure that data, data models, algorithms, governance structures, and technologies work in unison not just to enhance one business area or function, but to advance the entire organization forward.

Strategy is driven and crafted based on the ecosystem within which the organization operates. An ecosystem is a group of living, interactive, and interconnected elements that operate under explicit and implicit rules in a given environment. There is an ecosystem that underlies a data strategy. Understanding this ecosystem as it applies to each organization is critical. Like natural ecosystems, every ecosystem has a set of common elements, but even ecosystems that belong to a given category have their own unique features. In other words, the ecosystems for two firms in the financial services industry have certain common elements, as well as their own unique features and elements. Remember we discussed how organizational culture influences data governance

success? So, two similar firms in the financial services industry may need two entirely different strategies when it comes to data governance.

Align Data Strategy with Business Strategy

Data strategy is a subset of business strategy. Creating a data strategy without understanding or aligning it with the business strategy is like building two arbitrary-sized wheels for a bicycle. The bike won't go far, and the rider is likely to fall and get injured. There are many reasons why close alignment between business strategy and data strategy is critical.

First and foremost is the need for resources to implement data governance. Data governance initiatives can be an expensive proposition. Data governance is an ongoing initiative. No organization can declare that it has achieved its data governance goals and that it is ready to abandon its data governance initiatives. In other words, data governance is not a project with a start date and an end date. Second, data governance initiatives don't always end in a tangible product – something that everyone can see, touch, or experience. It is like the foundation of a building. We don't see it but without it, the structure and safety of the building are compromised. This makes the job of a data governance leader that much tougher – he or she must sell an idea that is critical, but not always tangible or visible. These two reasons alone make it essential to align the data strategy with the business strategy. Third, when data strategy is aligned with the business strategy, data governance leaders can clearly demonstrate how their contributions enhance the key performance indicators (KPIs) of an organization such as increased profits, reduced costs, less talent attrition, higher levels of innovation and productivity, enhanced brand reputation, market dominance and superior customer service, and reduced risk. Fourth, making data strategy a subset of the organization's business strategy opens the door for a more inclusive conversation around how to move the organization forward in terms of data maturity, monetization, and effective usage. In other words, the business units and IT can establish meaningful collaborations across different departments and units if improving organization-wide KPIs is the key focus of data governance. Subject matter experts (SMEs) from business functions, IT, business processes, data science, and project management can come together to provide rich perspectives on the data issues and challenges that they face. It helps individuals to think holistically about ways to work together to benefit their own area or unit, while advancing the organization. This is not to say that there will be no

turf wars or conflicts or hidden political agendas. It simply means that the odds of success are higher as everyone in the organization is challenged to take a more holistic view of data. It is a little harder to argue against why a KPI that is critical to senior management does not deserve the full benefit of a data governance strategy.

This evidence-based approach of showing how data governance has a direct impact on KPIs resonates profoundly with senior leaders across all units, departments, and functions. Unit leaders face their own unique data-related challenges that stem from multiple causes such as poor and untrustworthy data, data duplications, data that are not fit for use, inappropriate sharing of data or their opposite problem, not sharing relevant data, and many other related data challenges. By showing how their data-related challenges can be addressed through governance techniques that leverage the effective use and power of data, data governance leaders will get buy-in from senior leaders, without which data governance initiatives often die on the vine. The C-suite is willing to allocate resources when it sees a clear and categorical link between governance and the bottom line. More importantly, they become advocates of data governance within the organization. It no longer remains an abstract concept, a passing fad, or an IT issue. Instead, data governance becomes an essential tool to strengthen KPIs within the organization.

Mazzei (2017) identifies three approaches or ways to view data. Most organizations view data as a necessary tool, like money or physical assets. There is nothing wrong with this approach as long as it is anchored in a strong, well-understood, and dynamic data strategy. In the absence of a data strategy, viewing data as a tool often minimizes the power, legitimacy, ownership, and usability of data. When organizations have a meaningful strategy that drives data-driven decision-making, the potential to monetize data increases significantly ([Table 1](#)).

Table 1.Three Tiers of Value Creation.

Data as a tool	Managers are able to solve traditional value chain problems more efficiently and effectively; existing capabilities are improved through real-time, customized decision-making for individual consumers
Data as an industry	Spin-offs and new ventures are created to specialize in acquisition, storage, and analysis of data, construction of infrastructure, and development of software devoted to handling big data
Data as a strategy	Visionary leaders develop companies dedicated to building data resources to allow them to develop radically innovative business models that wed traditional and modern strategic thought

Examples

The best advice in the abstract can often lose value in the translation to the

tangible. Here we seek to provide concrete, albeit generic, examples of practices that might be followed in strategy development and alignment. Let's start by looking at the data counterparts of a fictitious business strategy^{xiii}:

O-Foods

O-Foods offers a line of organic foods that range from cereal and soups to cheese and chicken. These products are currently sold at three company stores, and those products that can be easily shipped are also sold through the O-Foods website. In addition to the O-Foods branded products, they also carry other organic products.

Each of the company stores is located in a town with a major university; both students and working adults interested in a healthy lifestyle are their target customers. O-Foods attracts and retains a talented staff that provides product information along with a high level of service to its customers – creating a strong competitive advantage.

The following strategy and goals were developed based on their SWOT analysis and three-year vision.

Strategy

O-Foods will provide its customers with the freshest and finest organic food available. O-Foods' team ensures each customer has an outstanding shopping experience by providing a high level of service and product information. O-Foods will:

- Implement an employee stock ownership program for those employees who have worked for the business for a minimum of one year.
- Develop an employee training program to further enhance all employees' product knowledge and customer service skills.
- Create a unified look for the stores, the website, and the product packaging.
- Design a marketing plan to increase purchases by current customers and attract new customers within the existing target markets.
- Increase operational efficiencies.
- Improve shipping capability so that perishable products can be ordered

from the website and shipped anywhere in the United States.

Goals

- Develop incentives to maintain high employee retention with a turnover rate of less than 20% annually for those at or above the assistant manager level and less than 30% for all other employees.
- Build the O-Foods brand.
- Increase profits by 15% each year for the next three years.

Given this information on the organization's strategy and goals, what might the elements of a data strategy look like?

- What data elements might support the training program? Will there be data on training activities that can be captured? What metrics might be developed to show program success? Identify these items in your data strategy and show how the data management or governance activities directly contribute to this strategic activity. Possible data strategy element(s): Define, collect, and manage data on employee training. Develop metrics to show employee progress in product knowledge and customer service skills.
- Data are at the core of marketing activities. What data management or governance activities can improve the capabilities of the marketing department or support the success of a marketing plan? These elements should appear in your data strategy. Possible data strategy element(s): Improve accuracy of data on current customers and potential leads. Incorporate external data sources to better target potential customers.
- How inefficient are operations currently? What data do O-Foods have to show the current baseline? What targets and metrics can be defined using operational data? Are there other data sources that need to be acquired? The data strategy should outline what data functions are available to support this strategic priority. Possible data strategy element(s): Create and manage time-to-market metrics for key product lines. Provide data and dashboards on key business processes to expose efficiency issues and highlight production problems.

- What data can be brought to bear for showing the achievement of the strategic goals? Ensure that the data strategy includes elements of how data can be used to highlight where goals are met and where a new strategic approach may be needed. Possible data strategy element(s): Create monthly data extracts for relevant departments (human resources, marketing, and finance) to support local reporting and analytics for measuring goal attainment. Create dashboards for KPIs to allow more real-time monitoring using operational data available.

Offensive versus Defensive Strategy^{xiv}

Dallemlue and Davenport (2017) make a strong case for understanding the nature of a data strategy. Is it defensive or offensive? How does one know what approach to take? It goes back to our earlier discussion about anchoring data strategy in business strategy. If the business strategy is defensive, the data strategy must dovetail this approach to a large extent, although pockets within the organization may take the opposite approach.

A defensive strategy focuses mostly on minimizing risk, while an offensive strategy focuses on leveraging data sources, application, and knowledge to achieve market dominance. A defensive data strategy focuses on ensuring compliance with regulations including data privacy, financial reporting, theft, and fraud and focusing heavily on a “single source of truth.” An offensive strategy may focus heavily on KPIs that help the organization to achieve market dominance, brand recognition, or higher industry ranking. It focuses on analytics to gain deeper insight into customers or members.

Defensive	Offensive
Not to get into trouble	Benefit from Data
Minimize risk	Maximize value
Compliance with regulations	Competitive advantage
Financial integrity	Increase revenue/profitability
Fraud detection	Customer insights
Single source of truth	Innovation and new product development
Legal, financial, compliance, and risk	Better decision-making

Regardless of the preferred strategy, the authors recommend selecting not more than eight priorities from the following list of organizational objectives as a foundation for developing a meaningful data strategy.

1. Create new products and services;

2. Improve revenue through cross-sell, pricing, and expanded customer base;
3. Meet industry regulatory requirements;
4. Develop analytics and digital capabilities;
5. Improve IT infrastructure and reduce data-related costs (number of databases, etc.);
6. Improve the quality of data and generate ROI in big data and analytics infrastructure;
7. Optimize existing strong bench of analysts and data scientists;
8. Mitigate operational risks such as data breaks, fraud, etc.;
9. Leverage new sources of data, internal or external;
10. Monetize the value of the company's data; use internal data as a product or service;
11. Prevent cyber attacks and data breaches;
12. Reduce general operating expenses and streamline business processes;
13. Rationalize multiple sources of the same data and information;
14. Use sophisticated, real-time, or near real-time analytics for business;
15. Respond rapidly to competitors and market changes.

We recommend that key drivers for data strategy be selected in close consultation with senior management. The KPIs of the organization should drive the objectives of data strategy. Buy-in from senior management is critical for the long-term success of data strategy. Aligning organizational KPIs with data strategy objectives provides a platform for data governance leaders and experts to build robust metrics to measure the return on investment (ROI). Objective metrics can prove how investments in data governance initiatives lead to strengthening KPIs, whether the strategy is defensive or offensive or both. We reiterate this point further in the next section.

It may be tempting to classify strategies into one bucket or the other, we

recommend against it. Resist the temptation to classify the entire organization's strategy as one or the other since some units or business functions may need a defensive strategy, whereas other units may need an offensive strategy. However, having a general sense as to the nature of the overarching data strategy of an organization can be extremely beneficial. Almost every company needs both. However, with limited resources, an organization has to decide how to distribute its resources between the two. This is not as easy as it may seem, nor is it fixed once determined. Sometimes the industry may determine which is more important. Health care, for example, may lean more toward defense than offense.

Asking the Right Questions

Many of us are not formally trained to ask the right questions on complex issues. We often hesitate to ask “too many questions.” But questions, good ones and plenty of them, are the key to understanding how users interact with data and their perceived value of organizational data. Practitioners in the field will unequivocally state that asking the right questions and plenty of them is the best way to unearth the data culture in an organization. And, understanding the observable and indiscernible data culture is the key to building and sustaining a successful data governance strategy. We have provided a long list of questions to ask. This is not to say that it is important to ask all the questions listed below or that these questions are even relevant to all organizations. The “right” questions will depend on the level of data maturity of the organization, the data culture, and a whole host of other unique factors. The purpose behind this list is to trigger the right thought processes that can effectively drive data governance. If a question sounds irrelevant or untimely, skip the question. Individuals may not have the answer to some questions and in some cases, may not even know where to look for an answer. This is not uncommon. Ocean swimmers don't know every ocean creature! Make a note and move on.

As work progresses, other questions that are not listed here will also surface. Keep a log of those questions and categorize them as appropriate by functional area, application, roles and responsibilities, application, or any other category that makes aligns well with organizational goals and objectives. It may make sense to categorize the questions by time. Some questions may become relevant in six months to a year, while others may be relevant right now. Questions trigger thought processes and reveal hidden problems, challenges, and opportunities and hence, play an important role. Below are a set of sample

questions.

Quality, Quantity, and Source of Data

- How much data does the organization currently collect? Own?
- What are the primary sources of data?
- What are the secondary sources of data?
- Are source data clearly documented and traceable?
- What percentage of the data that the organization creates or owns is used?
- What is the difference in the usage rate for structured vs unstructured data?
- What are considered critical data in the organization? Why?
- Is there universal agreement on what constitutes critical data? If not, why?

Cost

- Are there regular reviews of data-related economies-of-scale activities, such as creating, storing, accessing, and protecting critical data?
- What is the cost of creating and storing the data? Does the benefit outweigh the cost?
- What is the information architecture? When was it last updated?
- What is the data architecture? When was it last updated?
- What is the capacity and scalability of our infrastructure?

Ownership, Usage, and Accessibility

- Who uses what data? How frequently? For what purpose?

- To what extent are reports, algorithms, models, and other data analytical tools standardized, updated, and communicated to all constituents?
- What are the reasons behind low data usage, if it is the case?
- What are the data silos as they relate to KPIs?
- Are data ownership issues clearly outlined and enforced?

Risks

- What are the risks associated with critical organizational data?
- What are the weak links in the organization in terms of data security, integrity, usability, and accessibility?
- How much data are shared with external entities like partners, suppliers, associations, regulatory agencies, and the government? What are the risks and exposure involved in these sharing agreements?

Data Quality and Integrity

- Can business users access common data from a single trusted source rather than from different business applications?
- What, if any, are issues and challenges related to data licensing?

Communications

- Does the business speak a uniform data language?
- What percentage of time do decision-makers spend on locating data versus analyzing data to enhance the quality and timeliness of their decisions, particularly as it relates to their KPIs?
- Is there meaningful and regular communication and collaboration between projects, units, and departments about data issues, methods, protocols, and strategy?

- Is important data, such as information related to customer touchpoints, shared across critical core functions in a coordinated manner, such as marketing and sales, product development and innovation, and finance?
- Are data roles and responsibilities clearly delineated?
- Do individuals clearly understand their roles and responsibilities with regard to data, and if so, what are the performance metrics for those with data responsibilities?

Strategic

- Is there potential for monetizing the data? If so, how?
- What are the business KPIs and their relationship to critical data?
- What is the level of data maturity in my organization?
- Do we have the necessary talent to build a strong data culture?
- What is the level of stakeholder trust in organizational data? How does it vary from one unit to the next?
- How can decision-making related to KPIs be strengthened and improved with data?
- What data do employees currently need that they don't have because of lack of trust in existing data, not knowing where to find them data traceability issues, lack of data standardization, inappropriate formats, incomplete or inconsistent data sets, or accessibility issues?
- What metrics are used to assess data governance?

Roles and Responsibilities

- What are the different data roles in each unit? In the organization?
- What is the extent and depth of employee training for data governance activities of an organization?

Symptoms of a Weak or No Data Strategy

It is tempting to jump right into action and initiatives when it comes to data governance. There is so much work to be done, and the pressures on data governance leaders to get it done are great. Further, building a strategy is not easy work. It is a time-consuming activity that involves stakeholders at all levels in an organization, often with multiple or conflicting agendas at all levels in an organization. It requires not only building a great strategy but also getting buy-in, which can be time-consuming and frustrating. It requires building a business case and asking for resources from senior management who may not be completely convinced about the value of data governance. In spite of these challenges, moving forward without a data strategy is not an option. Organizations that charge ahead without a strategy can face debilitating long-term challenges that can cause serious setbacks in terms of increased risk, loss of revenue, brand damage, customer and talent attrition, and noncompliance with internal and government regulations. The telltale signs of no strategy or poor strategy or poorly executed strategy are many, a few of which are listed below:

1. **Ignorance about customers and their needs:** Both for-profit and nonprofit organizations have an obligation to their customers and/or members. An intimate understanding of customers is the key to long-term success. Organizations without a data strategy have little to no understanding of their customers, their needs, behavior and buying patterns, motivations, product-profit ratio, and market dynamics that triggers positive action on the part of the customer. Such organizations may not even have an accurate record of each customer. They usually don't have a universal organizational language that showcases a customer in the same light throughout the organization. The consequences of lack of clear definition (refer to Chapter 2) lead to frustrated employees, disappointed customers, lost revenue and market opportunities, and brand damage.
2. **Obstacles to effective decision-making:** Employees face insurmountable challenges in finding the right data in the right format at the right time. Even when they do, they sometimes don't trust the source of the data. Duplicate records are confusing and inconsistent. Decision-makers spend an extraordinary amount of time locating data, cleaning

them up, and checking their integrity. Decisions get delayed, simply don't get made, or result in wrong and untimely decisions. In some cases, this lack of trust in the data can lead to ignoring even good data. In other words, the organization limps along without leveraging the data it has to achieve market prominence.

3. **Conflict and Blame Game:** Without a clear data strategy that outlines the roles and responsibilities of key stakeholders and provides great clarity around the issue of ownership, there is perpetual conflict between IT and the business units. Confusion and chaos lead to finger-pointing while critical issues never get resolved. Project managers create their own systems, algorithms, and protocols to navigate such conflicts, resulting in wasted resources and duplication of efforts. In some environments, IT and business units go as far as to sabotage the efforts of the other. These conflicts result in poor ROI on IT investments resulting in the C-suite questioning the value and viability of all IT investments.
4. **Disengaged C-suite:** A meaningful and dynamic data strategy that produces results that directly impacts KPIs will engage the C-suite. The opposite is also true: senior leadership that is ignorant of the power of data or that does not understand the challenges of implementing successful data governance initiatives will remain disengaged, further amplifying existing challenges and headaches. Without a clear data strategy that outlines responsibilities, and data accountability and ownership at all levels in the organization, the C-suite itself becomes a problem riddled with conflicting agendas, disengagement, and ignorance.
5. **Level of Risk Increases:** Data have a close and intimate relationship with organizational risk. Data integrity, accessibility, reporting, sharing, usage, and compliance impact risk. When data are not closely monitored and are unattended to, risk creeps in from unexpected and unmonitored sources. Organizations rarely have the resources to completely eliminate all risk, but all organizations have the responsibility to reduce their risk to the fullest extent possible within existing resource constraints. Without a data strategy, the component of risk increases and can lead to disastrous results.

Conclusion

Data strategy is difficult and challenging. There is no manual as strategy depends on the culture of the organization and its data maturity level, among other things. While it may be tempting to craft a day-to-day strategy depending on the challenges of the hour, such an approach has long-term negative consequences. Hence, even though the task of crafting a meaningful strategy is time-consuming, demands managing and negotiating conflicting agendas, and involves active engagement with key stakeholders, many of whom may not understand the importance of data governance, data strategy is foundational to organizational success. The ability to ask the right questions and to doggedly resolve and pursue answers to the data challenges is critical to long-term success and sustainability.

Questions to Consider

Why Data Aren't the New Oil

Part of the motivation for writing this book was to help fill in the blanks where we found that the wisdom of the data governance gurus seemed disconnected from the experience of working through data governance issues in an environment where governance was sporadic, uncoordinated, and not well understood. While much of the advice is sound and well thought out, in many cases it is not useful, especially for new practitioners. We've come to realize that it may be because some of the advice doesn't necessarily reflect the reality that many organizations face. Let's take the "you can't govern what you don't know you have" adage. It is often the opening argument for why a data inventory is necessary and is often used in conjunction with the "Data are an asset and should be catalogued and managed like one" argument. While we completely agree with the sentiment, often the advice and guidance on how to achieve this lofty goal relies on analogies with physical assets, like office furniture, or other valuable commodities, like oil. The implication is that such an undertaking should be straightforward. However, nothing is further from the reality that many organizations experience. The trouble with these comparisons as a useful basis for helping organizations think about creating a data inventory is that most tangible assets have an agreed-upon definition and metric for

counting the single physical instance of the asset. Data have none of the above.

If we torture the “data are the new oil” analogy, one can argue that institutions know that the oil they want to count is a petroleum product. There may be disagreement over whether to count Brent Crude or West Texas Intermediate, but no one would question whether olive oil should be counted. Not so with data. What data are being counting? Things that are primarily numeric and structured like spread sheets? Databases? Reports? Unstructured content like text? Anything born digital? As mentioned previously, there needs to be agreement on what data are before they can be counted. Not only is there not industry-wide agreement, it is often difficult to get agreement even within a single institution. There have been countless enterprise data inventory efforts where even the group charged with finding and documenting “data” couldn't agree on what to include. Some participants only want raw inputs or data that were created as part of a business process. Others want to include the analytical outputs, including reports and visualizations, as data. Who is to say which is correct? A more appropriate adage for discussing data inventories may be “if you can't define it, you can't count it.”

Now let's talk about the metric for counting the asset. Oil is counted as a commodity collected in 42 gallon barrels. That is the standard. It's an odd standard that grew out of a historical transportation issue – transportation of oil in the nineteenth century was subject to leakage and 40 gallons needed to be delivered, so 42 gallons were used to ensure the requisite amount reached the final destination.

It may be old fashioned, but nonetheless, it is a standard. What is the “barrel” for data? What constitutes a data object to be counted? What is the vocabulary to be able to answer that question? This of course is inherently tied to the “what are data” question, but even if we say that we just want to count numeric data that are readily formatted in rows (observations) and columns (variables) like a spreadsheet, how are those counted? Is it each tab in the workbook? The entire workbook? What is the vocabulary used to describe the objects? A data collection? A data set? Often the vocabulary is borrowed from technology: a database table or view, a CSV or Excel file. But those words describe the barrel, not the oil inside. Again, the language is important but the metric is key: “if you can't define it, you can't measure it.”

And finally, there are often articles and posts that posit the questions about why businesses can count their office furniture and not their data. It's fairly straightforward: data are digital and there may be no single manifestation of a digital object! One can count barrels of oil or office chairs because they are tangible items with a single representation. If an organization has a barrel of oil and moves it from one place to another, it can easily be tracked because it is a single physical instance of the object and it can't exist in two places at the same time. If an organization has some data, they can make a copy of them and then have the original version in one place and the copy in another. Is that one item or two? The content is exactly the same in that it is the same logical data object but there are different physical instances of it. Conceptually, same content. Physically, different files. What if the content is slightly different? If I make a change to the copy, does that mean it should be counted as a different data object? How much of a change? How can anyone know it's enough to make the copy “different”?

It is often posited that “copies are bad” and “there is only one single source of truth” following questions like this. Except that copies are not inherently bad. There are very good instances where multiple versions of the same content are necessary. A transactional system that contains the data to actually run a business process may not be where reporting should be done. That is usually the reason for creating a data warehouse, but an organization may not want to do advanced analytics there either. The advice is that data lakes are where the data scientists swim. How do transactional repositories, data warehouses, and data lakes relate to the concept of a “single source of truth”? That may not be a useful concept for many organizations because it may not exist and it may not make sense. It may be more useful to discuss “authoritative sources” and “trusted sources” because those may make more sense. And there can be more than one. And good data governance practices would be to define them and the data objects for which they are the source. And then count them for an inventory.

1. What is your organization's business strategy? Where can you find relevant data touchpoints?
2. How often does strategic planning occur? What opportunities are there

for data-oriented conversations as part of that process?

3. Do you see signs of a weak data strategy in your organization? Which ones? How might you propose to correct that?
4. What questions do you think will elicit the most relevant information for a data strategy in your organization? Why? What might be different in a year?

5

Data Governance Frameworks

We repeat what we have stated in earlier chapters: Data governance is a daunting task. We state this over and over again not to deter those with data governance responsibilities or to limit the enthusiasm of the C-suite to invest in data governance. Our intention is to convince decision-makers that the rewards of data governance far outweigh the challenges it presents. Most senior leaders would be more than willing to invest in a tool that has a direct positive impact on the bottom line, enhances the productivity of employees' multiple fold, offers rich insights into market opportunities, and improves customer service. Data governance is one such tool and data governance frameworks provide a structured approach and processes to advance the value of data as a corporate asset.

A framework is a set of processes designed to ensure that data and information assets are formally managed throughout the enterprise. Frameworks must ensure that all aspects of data governance are addressed: stewardship, information governance, data definition and usage standards, master data management, metadata management, data life-cycle management, risk, and cost containment. They should outline the structures, workflows, and participants that will be part of the data governance program. A data governance framework is essential because there are two wheels that must move forward simultaneously: the organization must continue to make effective decisions on a daily basis even as data are pouring in. The wheels of effective decision-making must keep up with the wheel of managing the integrity and accessibility of a relentless flood of data and information.

A meaningful, yet flexible, framework that accommodates the shifting data needs and compliance regulations of an organization can provide lasting and extraordinary benefits, a few of which are listed below:

1. Provide clarity regarding the mission of data as an asset to each stakeholder.
2. Increase the confidence in organizational data, and thereby leverage their use for better results (Delacroix & Lawrence, 2019).
3. Articulate clearly the roles and responsibilities of each stakeholder; outline the portfolio of accountabilities to minimize finger-pointing and blame (Kamioka, Luo, & Tapanainen, 2016).
4. Prioritize what is strategic and important, even as the organizations face constant change.
5. Define in measurable terms metrics related to data and establish a clear roadmap to monitor and achieve success.
6. Facilitate meeting and/or exceeding compliance and regulatory goals and requirements.
7. Create better structures for timely retrieval and retention of data, thereby leading to better decision-making and stronger relationships with customers and internal and external stakeholders.

Data Governance Structure

One of the fundamental aspects of any governance activity is to understand who is doing, or is expected to do, what with the data that an organization is trying to govern. The corollary to that understanding involves setting rules, policies, or guidelines about who should or should not be doing what with data being governed. Identifying and defining roles and managing expectations about those roles is an important early step in getting a data governance program up and running.

Data roles and their labels were outlined and discussed in Chapter 3. Now we focus on what mixture of roles might be helpful and how those roles can be organized to make up the data governance structure. In general, most governance structures have at least three layers:

1. **Executive/strategic:** Approves initiatives and funding requests and endorses outcomes and recommendations. Participants at this level are

expected to set strategy, allocate resources, and provide final approval for processes and policies.

2. **Tactical:** Contributes to initiatives approved by executive group and makes recommendations for solutions to identified and prioritized issues. Participants at this level are expected to implement strategies, coordinate work, and suggest solutions and possible process or policy changes.
3. **Operational:** Implements suggested processes, follows policy, and raises issues to tactical and/or strategic layers. Participants at this level are performing the relevant data management tasks which are directed by data governance policies and processes.

The organization of staff within these layers and the shape of the governance structure will be heavily culture dependent. As discussed in Chapter 2, organizational culture shapes all activities whether intentional or otherwise. Designing anything that does not reflect the expectations of the participants is an invitation to failure. Some organizations will want to have different names for the roles at different levels; the executive level may have the “data trustees” with the “data stewards” and “data custodians” at the tactical level and the “data users” at the operational level. Other organizations may decide that using the “steward” label is appropriate for all levels and therefore have “executive” or “domain” stewards over the “business” data stewards who are a level above the “operational” data stewards.

Beyond labels, the scope and reach of each layer may be different. For organizations with a strong hierarchical structure, the structure of a data governance framework is likely to be a small executive layer that is likely to be over a larger tactical layer with the operational layer having the largest number of participants. Such structures are usually depicted as a triangle as shown in [Fig. 4](#). Often there are support services or other cross-layer options portrayed as structures along the sides of the pyramid. These organizations are likely to have the hierarchy-oriented, or power, culture described in Chapter 2.



Fig. 4. Hierarchical Data Governance Structure.

As the degree of hierarchy flattens or decreases, or degree of cultural collaboration increases, the degree of difference in width between top and bottom that is appropriate lessens and the resulting structure is more likely to be rectangular. The layers are still the same or similar but organizational cultures built around consensus and collaboration tend not to have strong hierarchies in practice, like the Clan culture or Task culture. Therefore, there may be a larger number of strategic stakeholders relative to the tactical and operational participants. Such a structure may look like [Fig. 5](#). In addition, large or especially decentralized organizations need to consider whether or not one structure is appropriate. For some entities, a single centralized group charged with making strategic decisions for the entire enterprise is just not feasible. For others, it may not be palatable. In either case, a successful framework will

accommodate the structure appropriate for the organization. It may be that different divisions, subsidiaries, departments, or other subsets of the organization each has its own data governance structure that matches the culture and responsibilities for that area. Then a representative of that structure can be part of a broader steering committee or other group that focuses on the enterprise.



Fig. 5. Collaborative Data Governance Structure.

Deciding which areas participate and how can help an organization determine the structure that matches their goals. For an organization focused on data governance to achieve regulatory or compliance goals, representation from the legal, audit, and information security areas will be critical and may come with the authority and responsibility to lead policy discussions. For organizations more focused on improving data sharing and appropriate usage, representation from domain and business experts may hold more sway in strategic discussions. Refer to the cultural chapter for the impact of expertise on decision-making. Hierarchical organizations may easily be able to identify the critical areas that

should be included but other organizations may want to include a broader range of participants to help with adoption and acceptance.

Chief Data Officer

Regardless of the shape of the structure, one of the really important decisions that needs to be addressed is who will be the point person for data governance and how will their role be defined. This discussion often devolves into the “Do we need a Chief Data Officer (CDO)?” conversation. Regardless of the title, a central role focused on data issues in general, and data governance issues in particular, can be incredibly helpful, especially for organizations just starting on a data governance journey. First, a CDO or equivalent has a full-time data focus. Unlike roles in other areas, say marketing or finance, data are not a means to an end but are the end in themselves. Having someone who is thinking about data all the time and keeping track of the variety of data-related issues that arise across the enterprise is valuable and it enables a true organization-wide data governance program to be more effective. Second, it is important to have someone with a strategic view who can articulate the issues and value proposition for data governance and data management activities to both the C-suite and the staff. Knowing where the organization is headed, or would like to be headed, can inform the data strategies and help coordinate activities across the various silos. Third, this individual can serve as an effective connector across even tactical data operations that might be housed in different silos. Being charged with enterprise-wide data governance can open the door to cross-functional conversations in a way that is more effective than the ubiquitous committee or workgroup solutions that many organizations try. Finally, an individual with a concentrated data focus may be a more successful evangelist for good data practices and collaboration on important data initiatives. Besides being the “single throat to choke” that some people may want for accountability, they can also be the purposeful cheerleader to help lead the change efforts that can be so difficult as we outlined in Chapter 2.

All these reasons center on a crucial role for any person charged with leading data governance: communication. As we discussed in Chapter 3, communication is so critical that having a centralized role responsible for data governance can help focus the messaging and provide consistent verbiage and direction for all the discussions, presentations, and meetings that are required for data governance to be successful. While it is true that a dedicated executive focused

on data governance isn't strictly necessary, it can make things much easier and given how hard data governance is already, we think it's worth pointing out anything an organization can do to make things easier.

Organizational Placement

The CDO role, or equivalent, provides an excellent launching point for having the important discussion: from where should data governance be led? As we have already argued, data are not an IT issue even though that is a common misperception. Yet, many organizations that decide to hire a CDO place that role within the IT department. Nearly a quarter of the CDOs surveyed by Gartner report to either the chief information officer (CIO) or the chief technology officer. In many organizations, the two roles are at worst conflated or at best specifically combined. Regardless, delineating between what is the purview of the CIO role relative to what is expected from a data governance program is critical for success no matter where the program resides or who leads it.

For many organizations, governance around information technology decisions, implementations, and spending are not new. Where such infrastructure and discipline are well established, a natural step would be to include data governance issues and discussions as part of the IT governance process or program. Such a solution will be expedient but may prove problematic in the long run. As we have discussed previously, data are not strictly an IT issue. While there are many instances where data and technology processes and decisions will go hand in hand, there are too many instances where that is not the case for such a marriage to be fully successful.

One of the fundamental differences is on the focus of each program. Data governance will be emphasizing the “what” side of data: What is the definition of customer? What are the rules for sharing data? IT governance, on the other hand, is focused on the “how” side of data: How do we store customer information? How do we share that information? Sometimes the policies, processes, and people involved in these decisions are the same but often they are not. For organizations with established IT governance practices and a nascent data governance program, it will be natural to model a data governance framework after an IT governance framework. Such an approach often yields very good results. When data governance is specifically placed within the technology area of an organization, it becomes more difficult to practically separate the two in terms of participants and decisions to be made, and the

organization often winds up using one framework or program to accomplish both. This design tends to continue to perpetuate the belief that data are a technology issue which can frustrate both business and IT staff if roles and responsibilities for both data and technology decisions are not clearly and strategically defined.

The emerging view for many organizations is that if data are a business issue, then responsibility for data governance should reside with the business side of the organization. In some instances, this means that data governance is the realm of the chief financial officer or chief risk officer, arguably roles with the most stake in data sharing and data quality policies and processes. In other organizations, data governance is driven by the chief marketing officer who may be best placed to see how data governance can provide the most potential for revenue generation (Laranjeiro, Soydemir, & Bernardino, 2015). The most common non-IT placement for data governance activities, however, seems to be in a standalone function reporting to the CEO or COO. For many organizations, especially highly siloed ones, a separate group charged with working across all business lines provides the least objectionable placement. If there is a tight and strong tribal nature to the silos, placing another C-level executive within an existing silo may cause friction between various senior leaders who may see this as a power shift. Even without a CDO, data governance activities emanating from an existing business line, including IT, may be seen as being irrelevant to other lines of business. More and more often, consulting firms and other data governance organizations are recommending the standalone placement as providing the best opportunity for success for a variety of corporate cultures.^{xv}

Research versus Institutional Data

Throughout this book, there has been an unspoken assumption that the data in question for a data governance program are the institutional data which organizations use to do business. Many entities only have this kind of data so such a focus is reasonable. However, many organizations may have a research department or R&D subsidiary who are using data to go beyond the typical business questions of how to improve operations, reduce cost, mitigate risk, or whatever motivation is behind the collection and use of internal data. Research data are different than institutional or business data and require a different kind of data governance framework. Whether at a corporate entity or an academic

institution, research data still need to be governed but not the same way that institutional data are. This difference isn't always well understood so we look to explain it further.

We invest and conduct research to understand the world around us. For some disciplines, that means understanding the economy. For others, it might mean understanding an ecosystem or solar system, a biological function, or human behavior. Research doesn't have to use data, but more and more often it does. And those data need to be collected, created, or procured in order to be analyzed. Here we use the following definition: Research activities use data to answer a question to better understand how the world works.

Not all analytics are research. Organizations in a variety of industries also use data for analytics. But those data are nearly always about the business, and the questions that they are trying to answer involve improving the business. Sometimes the analytical task requires additional external data in order to answer a particular question, but the subject of the question is nearly always internal. Here we would use a second definition: Enterprise activities, such as business intelligence, use data to answer a question to better understand enterprise functions and improve business processes.

Point: Research is outwardly focused; enterprise analytics are inwardly focused.

Using these definitions, a second distinction comes to light. Research activities seek to answer a question about the world around us. Questions, once answered, may never be asked again. It is true that research projects tend to build on previous research projects. But such extensions generally focus on a different aspect of the question and do not ask exactly the same question in exactly the same way again. In addition, answers to questions might not be actionable by those who asked them. If a researcher learns something new about how the world works, they may not be the ones in a position to use that information to make a change. Understanding what causes cancer or how a solar system is born does not necessarily allow the researcher to prevent cancer or explore a solar system.

Enterprise activities, on the other hand, are usually repeated and in many cases are designed to be so. Regular monitoring of enterprise functions or business processes are often crucial to business success. Dashboards and metrics are the lifeblood of performance measurement and are used to answer the same question in exactly the same way by design. In addition, the answers to the questions can be used to make changes by the people that asked the question. A

key performance indicator that shows a process failure can lead the analyst that is responsible for its creation to initiate changes to that process.

Point: Research is generally done once without an expectation for action; enterprise is done repeatedly in order to take action.

A metaphor might be useful here. Let's consider the purchase of a new chair. There are two general types of chairs that one might consider.

One option is to purchase custom furniture: a chair that is hand built using special tools for a single purpose by someone with a specific skill. Being able to produce this same piece over and over is not the point of the process, but it can be reproduced using the exact same tools and materials if necessary. If it is well built, it will last a lifetime. It takes a long time to build and the chair may not be to everyone's taste. [Figure 6](#) shows an example of such an artisan creation.



Source: <https://www.flickr.com/photos/11932978@N00/4958243137>.

Fig. 6. Example of a Unique (“Custom Built”) Chair.

On the other hand, manufactured furniture is also an option: It is designed to repeatedly produce high-quality output. Each piece needs to be the same so standard tools are critical. Infrastructure is sturdier so that the production process can run regularly to meet production schedules and quotas. Production is streamlined and efficient to make the most people happy. Figure 7 shows a collection of matching chairs.



Source: <https://www.flickr.com/photos/flooded/8516605642>.

Fig. 7. Example of Matching (“Manufactured”) Chairs.

What Does This Mean for Data Governance?

Research data need to be collected, captured, created, or contracted from the environment being studied. Many research projects start with a research data management plan to outline how a researcher will collect, document, organize, manage, and preserve the data at various stages of the project. Such plans often include the following information^{xvi}:

1. The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

2. The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. Policies and provisions for reuse, redistribution, and the production of derivatives.

The goal is to allow others to understand the data used for the research and allow for potential replication or extension of the work.

Enterprise data, on the other hand, exist in applications or internal repositories and might need to be combined with other enterprise data to study a particular function or business process. These data exist to run a business function, and enterprise data management roles are important for ensuring that enterprise data are suited for answering business questions. There are well-established best practices and maturity levels, such as those described in Chapter 7 to describe how enterprise data are and should be managed.

For institutions that have both enterprise analytics and research functions, the dichotomy can be painful for data management and data governance solutions. Many data management practices that are essential for good enterprise data management have no relevance to research data management. Take data quality, for example. Good data management practices would outline how to define data elements, measure their accuracy, and report on the measurements, outline remediation strategies and decision rights for deviations from an agreed quality standard.^{xvii} For customer data in a marketing database or employee data in a personnel system, all of these activities are appropriate and important. For research data collected from a lab experiment or human subject interviews or possibly purchased from a third-party data vendor, none of these practices are relevant. The definition of the data elements may be dictated by the vendor or defined by the system or lab equipment. Accuracy may not be relevant or easily measured. More likely, it can be measured, but deviations from some standard may not be able to be addressed. There may be no decision rights relevant to changing an observation that comes from a survey respondent or a data broker.

Other aspects of enterprise data governance suffer from the same disconnect when applied to research data. Yet it can be difficult for organizations to

understand how to make allowances for such a dichotomy. That is not to say that research data do not require careful management, they most certainly do. But that management looks very different than data management for enterprise data. As analytics advance and more and different data are used in more atypical ways, understanding the difference and working to manage both types appropriately will be even more critical. Yet it seems that the enterprise data and research data worlds are largely unaware of, or possibly just unconcerned with, each other's existence. When designing a data governance framework, it is important to be intentional about how research data need to be addressed. It might be that a separate but similar structure to oversee research data issues can be established with formal ties to the enterprise data governance structure to ensure communication and commonality where relevant.

Cloud Computing Framework

While there are many benefits to cloud computing, which explains its popularity, from a data governance perspective cloud computing presents several challenges. First, control of data no longer resides within the organization but is transferred to the cloud provider. This is not an easy challenge to manage as the risk associated with data, including compliance, security, integrity, and accessibility, increases multifold. Regardless of what a cloud provider may promise, the responsibility of identifying, predicting, and managing these risks fall squarely with the data governance team. What is the best framework for cloud-based organizations? Develop a cloud-driven data strategy *before* moving to the cloud. Data governance for the cloud brings its own unique challenges and hence, a clear set of policies and procedures along with a clear demarcation of the roles and responsibilities of cloud vendors versus the organization must be in place prior to moving to a cloud environment (Al-Ruithe, Benkhelifa, & Hameed, 2016).

There are four common challenges that data governance experts in cloud environments face. While these are not unique to cloud environments, they can easily exacerbate the data headaches an organization already faces.

1. Assuming that data governance in a cloud environment is not any different from a noncloud environment. Lack of training and a strong communication plan with key stakeholders, both external and internal.

2. Lack of commitment on the part of the C-suite of the enormous challenges of implementing data governance in a cloud environment can lead to lack of financial support, poor monitoring, and enforcement regulatory compliance.
3. Standard data governance policies, procedures, and processes must be tweaked to accommodate the cloud environment, along with customizing the roles and responsibilities of key stakeholders to match the cloud vendor's offerings and services. Assuming that the cloud is no different from a noncloud environment can lead to problems down the line.
4. Managing a cloud environment requires in-depth technical knowledge and hence, aligning the needs of cloud governance with the organization's talent pool is essential. In the absence of the appropriate talent, cloud environments can create problems that can have a domino effect.

Big Data Management Frameworks

Although Big Data is now part of common business terms, it means different things to different individuals. Big Data is often described using several Vs – usually including Volume, Variety, Veracity, and Velocity, but there may be as many as seven so this is another area where there is no clear-cut definition. A practical definition might be that data are “big” if they can't be stored and manipulated in a spreadsheet. Not a perfect definition but for many organizations, once a data collection is no longer manageable using standard office tools, then special technology and business processes are required.

Further, the many subsets of data and information that emerge from under the “Big Data” umbrella also lack clarity. What we know and experience is that today we can process large volumes of data and study relationships that once were impossible to do. A second possible definition might be based on the structure not the size: “Big Data” are often the collections that result from combining multiple data sources. The ubiquity of data and the relative ease with which new technologies can be deployed to handle data outside traditional tools means that organizations are often looking to new and exciting ways to create larger datasets by combining multiple internal data sources or adding external

data sources. Regardless of the origins or motivations, large volumes of data create unique challenges for data governance.

The more data and data fields that organizations have to deal with, the greater the need for meaningful frameworks to manage Big Data. Based on the work of Tekiner and Keane, we recommend paying close attention to the following three stages as a suggested framework:

1. **Data Sources:** Big Data does not appear magically – it is the result of integration of large volumes of structured and unstructured data from multiple sources that are landing on the shores of an organization at speeds that are hard to keep up with. It is essential to ensure that appropriate rules for metadata and rigorous filtering processes are applied before integrating various data sources that make up Big Data.
2. **Analytics:** Computational power and sophisticated analytical tools are used on the Big Data collected in the previous step. Since Big Data uses internal and external datasets it is essential to ensure the quality and integrity in the entire data cycle of data from all sources. This is the role and responsibility of the data stewards.
3. **Decision-making:** The results of the analytics are then mapped to a decision model within the organization. New insights are derived from this process. Often, organizations are unable to derive the full benefits of this step because the volume of data, combined with new applications, new and emerging analytics, can cause more confusion or lead to wrong conclusions.

Big Data is the end result of interfacing with a number of internal and external hardware, software, databases, sensors, information technologies, business intelligence tools, and a large number of applications. Each of these interfaces creates a level of vulnerability for data governance advocates. Hence, critical expertise in these areas is needed if data governance initiatives are to be successful.

There are some special considerations for Big Data that we should point out here. First, data quality is still relevant and may require a different approach for data management. There are occasional references in popular press or quips made during presentations that with “Big Data” we no longer have to worry about data quality because we have so much data that outliers or incorrect

observations will be swamped by the rest of the data and will not affect any analytics. While it is true that in statistics, larger sample sizes tend to lead to smaller standard errors, one should not extrapolate this fact to imply that data quality is no longer a concern. In fact, the potential for data quality issues may increase with data that are joined from multiple sources if care isn't taken to ensure the merge is done appropriately. If there is any accuracy issues with the identifier used to join two datasets, the resulting record may contain a bad match. For example, if there is any uncertainty about whether customer John Q. Public is the same person as marketing contact John Q. Public, the resulting joined record may cause problems for the business units that are trying to simultaneously maintain a good relationship with an existing customer while another business area tries to build a relationship with the same customer.

For “Big Data” to be used effectively, the data management best practices must be used for the compiled dataset as they would be employed for the existing source datasets. This situation is complicated when trying to combine internal and external data. Whether it be government data or open data acquired without payment, or data purchased from the increasing surfeit of data vendors, incorporating external data into internal business processes requires a solid understanding of both sets of information. Do they have the same unit of observation? Do time periods match? Where are assumptions being made in order for the joining to work? While it is tempting to just “throw all the data into a lake and let the data scientists swim,” it is still important from a governance perspective to be able to know how to evaluate the analytics produced from the exercise, especially where they will be used to inform decision-makers.

Another governance concern for “Big Data” is the decision rights. Identifying stewards for data in source systems can be challenging enough but when those data are pulled out of the functional area they support and combined with other data to create a new data source, it can be even more difficult to know who is allowed to say yes, or no, to what uses of the data. If someone has access to, say, customer data and access to social media data, does that mean they automatically are granted access to a dataset created by merging the two? Or is there a different process by which such access is approved? What if someone has permission to use one of the source datasets but not another and they are combined? None of these questions are unanswerable but for many organizations, especially those just starting their data governance journey, they pose a level of complexity that can be difficult to manage without explicit consideration.

Conclusion

As we have mentioned several times throughout the book, there is no one-size-fits-all framework for putting data governance in place. Organizational leaders need to decide what they want data governance to accomplish, who will be tasked with doing that, and how it will fit into their culture before they can effectively design a framework for their data governance program. The close relationship between data and technology needs to be addressed head-on to ensure that the right conversations are happening with the right people to achieve the right outcome. The variety of data types, the size, and location of data storage are important considerations in today's data and technology environment and they should be able to fit within a broader data governance framework.

Questions to Consider

1. What kind of data does your department or organization have? Are there differences in the way they are or should be governed? Why?
2. What kind of data governance structure does your organization have or should have? How does it match to your corporate culture?
3. Who can or should be responsible for data governance? Why does that make sense? What objections or risks might that raise?
4. How would you describe the relationship between data and technology in general? How is that reflected in your organization?

6

Data Governance Components: Data Quality, Literacy, and Ethics

Introduction

Data governance touches on all aspects of data collection, management, and use across all parts of the data life cycle. Such broad scope can make starting a data governance program a daunting task. As we discuss throughout the book, not every data governance issue can or should be addressed at the same time or with the same intensity. This chapter looks to outline three components of data governance, namely data quality, data literacy, and data ethics, that are foundational and applicable across a broad range of data activities but are not often well understood.

Data quality is commonly used as the precipitating issue that brings data governance to the fore: “We need high quality data,” or a variant thereof, may be a common complaint. But understanding what “high quality” means and how to quantify it is not always straightforward. Data literacy, like literacy in general, is a basic skill to develop and define how people work with data. It has been identified by Gartner^{xviii} as one of the top challenges facing organizations today. And finally, data ethics make many headlines but may not be well understood in the context of governing institutional data. We’ll address each of these components in turn.

Data Quality

For many organizations, data quality issues are the primary driver for instituting data governance initiatives. Indeed, many trade publications and consulting

companies give the impression that data quality is the most important thing that a data governance program can achieve. But what exactly do they mean by data quality? Many organizations treat it like the old saying about obscenity: they can't define it but they'll know it when they see it. Like the discussion in Chapter 1 about the definition of data governance, there are a variety of possible definitions, and it is important that organizations define what data quality means for them. Many of the definitions involve some notion of “fit for use”^{xix} or satisfying “the requirements of the intended use” (Han, Kamber, & Pei, 2011) without any discussion of what that actually means or guidance on how to document those requirements. Unfortunately, this is not helpful for those just starting on a data governance journey or who might want metrics to quantify data quality levels.

We'll try to fill this gap by proposing a concrete way to consider data quality dimensions.

First, let's dispel a misconception: data quality is not just about accuracy of observations. Many people don't define data quality because it is assumed that everyone understands that high-quality data means accurate data. As discussed in Chapter 3, such assumptions about the commonality of definitions are problematic and lead to confusion. While accuracy is a key dimension of data quality, it is not the only one. There are a variety of considerations, some of which are context dependent and the suggested collection of dimensions is not universally agreed upon.^{xx} We posit that the following are key dimensions that should be considered when assessing data quality.

- **Completeness:** The degree of comprehensiveness of the data or how much data have been collected compared to how much data exist.
- **Accuracy:** The degree to which the observation matches the truth, correctness.
- **Consistency:** The degree to which the definition and representation of the data are the same for all observations.
- **Timeliness:** The degree to which the data are both representative of current conditions and available for use. Some definitions separate the first part into a separate dimension of “currency” and focus solely on the timing of availability. Because the two are so often tightly linked, we choose to use a more comprehensive definition similar to that of the

Enterprise Data Management (EDM) council.^{xxi}

- **Uniqueness:** The degree to which no record or entry is recorded more than once.

There are other data attributes that can be associated with data quality such as validity (i.e., the degree to which an observation uses an acceptable representation) and coverage (i.e., the degree to which the data set captures the available records) and accessibility (i.e., the degree to which data are available for use). Indeed, some research has outlined upwards of 30 dimensions (Loshin, 2001). Given the complexity of getting started on a data quality management journey, we choose to focus on a smaller subset of critical dimensions. Resources for considering other data quality attributes or approaches are available in Chapter 11.

Even with this limited list of data quality attributes, organizations are likely to face challenges in achieving a perfect level for each attribute for all data. In fact, some attributes are not achievable at the same time. The nature of accuracy, for example, often requires that a great deal of time and attention be paid to ensuring that the representation of a concept in the data can be validated against some authoritative source or specified business rule. If an observation cannot be a validated, processes and procedures often known as “data cleansing” may be employed to improve the accuracy of the data element. Such corrective action can take time and may include additional data capture or extensive review and may lead to a decrease in the timeliness of either the representation or the availability of the data. We find that, in practice, there is a tradeoff between accuracy and timeliness: data that are immediately available are often “dirty” and may be riddled with errors, missing values, or inconsistencies. Data that are accurate have probably been cleaned and that process may take a long time. Organizations will have to decide which dimension takes priority for which data. It is likely that different data types or data used for different purposes will have different priorities. Data used for compliance or regulatory reporting, for example, may require a higher level of accuracy and so may not be the timeliest data available. Data used for exploratory analysis or to pilot a data science project may benefit from using the timeliest data available regardless of the accuracy of those data.

Because of such tradeoffs, we recommend that instead of focusing on the impossible notion of “high quality” data and instead focus on having “known

quality” for data. Because different attributes are likely to be important for different uses, a single notion of “high quality” cannot affectively capture the differences and can lead to unmet expectations and frustration. If data need to be highly accurate to be labeled “high quality,” then any activity using data that are timelier but less accurate is likely to face scrutiny and mistrust because it does not employ the “good data.” If organizations instead focus on more complete descriptions of data quality that cover the high-priority attributes, then data users will be able to know better what data are fit for their purpose and decision-makers can better understand the potential limitations of any analytics based on the data quality dimensions for the data used in the analysis.

Quality Assessment Frameworks

We would be remiss to recommend focusing on “known quality” without a description of how such a label might be created or applied. Entire books cover data quality assessment frameworks and are important sources of information for those looking to create formalized data quality strategies and infrastructures. For organizations that are at a lower level of data management maturity, it may be easier to have a simple rubric like that proposed by Becker (2019). Although Becker considers eight dimensions which aren't entirely in alignment with the focus here, the basic structure of the recommendations provides a helpful approach.

Let's take completeness as an example. Becker uses a similar definition for completeness: “the proportion of data that is stored against the potential for 100 percent of the data to be stored” (Becker, 2019). In many instances, there will be a strong inclination to create a metric that is based on the percentage of data available. While it is true that many software packages have the capability to check for missing data and even calculate a concrete number for the proportion of missing items, such a metric is likely to give a false impression of precision. Say a data source is 86% complete but it would be possible to get 90% complete with an investment of time or human resources, how should one evaluate that cost/benefit tradeoff? How much “better” is adding 4% more data relative to the investment required? In addition, the percentage description only evaluates sheer volume of numbers not the relevant importance of the information being captured. Any notion of quality for a data set needs to consider if the missing values are in a key measure, say “customer name”, or in a less critical value such as “second email address.” Missing data in one have a very different affect than

missing data in the other. Becker's suggested approach is much simpler: assess the relative completeness based on a five-point scale addressing the issues of concern for missing values:

Completeness

1. Most data records are incomplete, including critical information.
2. Many data records are incomplete, but critical information is present in most cases.
3. Some data records are incomplete, but all critical information is present.
4. A few data records are incomplete, but all critical information is present.
5. All records are complete.

For many people, such a scale seems very imprecise. What is the measure for “most” versus “many”? What are the “critical information” elements? We see this as an opportunity for flexibility and customization: every organization can create their own thresholds and assign the relevant items the “critical” label as fits their business purpose. The scale can be expanded or contracted to capture the relevant classifications and amount of distinction necessary, thereby moving the discussion to a more general “how much data do we really need” realm rather than comparing percentages.

Building such a scale for each of the relevant quality dimensions allows organizations to identify the relevant merits and challenges with any data collection and gives users the appropriate information to decide whether data are fit for their particular purpose. If a use case requires the most up-to-the-minute information regardless of accuracy, look for data that have a 5 for timeliness and understand what analytic issues will need to be addressed if the accuracy score is a 3. On the other hand, a use case that needs the most accurate data possible and can tolerate a lower level of completeness can use the relative scores there to ensure that users have the kind of data required. Such an approach can help inform quality improvement efforts as well as inform users as to the known quality of the data at hand.

Data Literacy

Like other topics in this book, we will start with the definition. What is data literacy? Many definitions involve the ability to derive value from data in a business context. We prefer to be more specific about what is involved rather than to rely on the interpretation of what it means to “add value.” We define data literacy as the ability to read, write, manage, and analyze data. Like the general concept of literacy, defined as the ability to read and write, data literacy involves being comfortable working with and talking about numeric representations of concepts. This is not to imply that all data are numeric, but understanding numeric data is a foundational ability that is essential for working with any kind of data. Data literacy is often confused with “statistical literacy,” a term generally used to mean the ability to understand the use of data and statistics in the world around us. Data literacy goes beyond statistical concepts, however, and involves being able to critically evaluate data sources and definitions as well as analytical methods and techniques.

Now that we have described what it is, it is logical to discuss who needs it. General literacy was once considered important for a certain class of people. As more printed materials became available, the idea that a broader swath of the population should be able to read took hold and literacy is now considered one of the markers of an advanced civilization. We see data literacy taking a similar path: once the realm of analysts and “quants,” the increase in the amount of data in both everyday life and the business world means that data literacy is now relevant for a larger set of people. Not every job or activity is data focused, but a large proportion of them now generate data or have measures and metrics that help define them. Fitness trackers create data about people's activities; programmable thermostats create data about energy use; web surfing creates data about shopping and communication habits. Analysts and data scientists may be the ones spending the majority of their time working with data, but nearly every person in any position is faced with the need to use and understand data.

For businesses, understanding the need for data literacy in the workforce outside the traditional data-focused roles can be challenging. Often it is confused with the need for a new technology or technical training. Like other aspects of the data world, technology is an important partner but should not be the main focus. Understanding data and how to use them is relevant regardless of platform or technical skills. There are many ways that improving data literacy across an organization can improve business outcomes regardless of technology choices. As we discussed previously, many organizations are trying to be “data driven” or have a “data-driven culture.” This strategic focus is so prevalent that dozens of

books have been written to help organizations achieve that goal. Research done by Wolff, Gooch, Caverio Montaner, Rashid, and Kortuem (2016) suggests four types of “data citizen” that we can translate to data literacy needs in business:

1. Readers need skills to interpret and understand data as this has become increasingly important in every aspect of a business. We suggest that this is the basic level of data skill that should be provided to all staff of any organization looking to be “data driven.”
2. Communicators are the data storytellers. Sometimes using words, sometimes in pictures, they make sense of data and make it easier for others to understand. These people have a higher level of data literacy and can help boost the literacy level of others in the organization.
3. Makers are those who need higher skills to combine data with strategy. They help figure out how to use data to solve problems or highlight problems that need to be solved. People with this level of literacy are often the champions for data and data skills and can help propel an organization toward a clearer data focus.
4. Scientists are still the valuable analytical roles that need to “combine strong technical data skills with communication skills and in-depth knowledge of the domain of the data.” Once thought of as unicorns or purple squirrels, data scientists are still in short supply but are more common than in the past. These are the folks that show what is possible with data that the makers can build into the strategy.

Now that we've defined it and decided who needs it, the next step is to figure out how to get it. While there is a growing movement focusing on teaching data literacy in schools, most organizations have an urgent need to acquire these skills now.

How to Use It?

Defining the issue and teaching people the skills are important first steps to data literacy. They are necessary but not sufficient conditions. Having the skill to do something will not yield a benefit without the right conditions to use that skill. Being trained as a carpenter isn't much use for those who don't have tools or

wood or an understanding of what furniture or structures are needed. Data literacy is much the same: users need to have data, somewhere to use it, and some idea of what to use it for. For data literacy to be a successful contributor to becoming a data-driven organization, we suggest the following are necessary^{xxii}.

- **Access:** Without data or data-related content, data literacy isn't helpful the way general literacy isn't useful without books or other written materials. Regardless of the level of literacy, access to data and information is important to allow data literacy to truly take hold.
- **Examples:** To build data literacy, people need to see what is possible. Why learn to read or write without some goal as to what can be achieved? The analogous argument can be made for data literacy: why bother learning about data without an idea of what the benefit is.
- **Infrastructure:** Like all skills, data literacy requires practice, so people need some place to work with data and gain or use their new skills. It could be as simple as a software license and shared file system or as complex as a data analytics platform but without somewhere to flex their data literacy muscles, any teachings learned or skills started will atrophy.
- **Critical thinking/filling skill gaps:** Even with data, and infrastructure to use it and examples to work from, data literacy will still need to address any skill gaps, either technical or critical thinking, in order to truly be able to transform an organization. Training is critical: It's not enough to provide libraries, books, and role models to the illiterate; we still need to teach them how to read. They need to be able to do it themselves, and the same is true for working with data.

Data Ethics

Data use and misuse seem to be in the headlines regularly. News about collecting data without consent, using data for things other than what was proposed, or using data incorrectly can cause suspicion and mistrust, especially between the data illiterate and the data savvy. There are many resources available to help guide organizations through the maze of ethical issues around “big data” and data science. We will focus on a topic that has stymied many data governance efforts: ethics and responsible use for internal data.

Like the issues currently making headlines, many of the issues for internal business data are focused on personal information. We do not want to give the impression that data about people are the only ones that require ethical consideration. Indeed, one definition of data ethics is far broader, describing a branch of ethics that “studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including artificial intelligence, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g., right conducts or right values)” (Floridi & Taddeo, 2016). We will touch on some of these topics; others are outside the scope of this book.

One additional clarification is necessary: data ethics is not strictly about privacy. While personal data have the tightest tie to privacy issues and are currently the focus of most of the privacy and protection laws on the books (e.g., HIPAA, FERPA, GDPR, etc.), privacy and ethics should not be confused or used interchangeably. Ethical use of data is a necessary but insufficient condition for protecting privacy. Many data activities that are ethically questionable do not violate privacy at all. In fact, some early data ethics issues arguments in research centered around the issue of whether someone's privacy was violated rather than whether or not the researcher did the right thing.^{xxiii} Discussions here will not focus on privacy because it is implied by ethical data activities. Instead, we look to outline some of the less clear ethical issues that organizations frequently face.

Two general adages can be applied here to most situations:

1. Just because you can, doesn't mean you should. Data users need to understand that because something is legal, that doesn't always make something right. There may not be a rule, law, or policy that prohibits some data activity, but it still may not be the right thing to do.
2. If you don't want anyone to find out about it, maybe you shouldn't do it. While it is true that there are ethical activities that may still need to be kept secret, especially around innovation and intellectual property issues, many instances involving obfuscation should be evaluated to ensure proper adherence to both the letter and the spirit of any laws, rules, or policies.

These aphorisms can be evaluated across three major data activities:

collection, sharing, and use.

Ethics and Data Collection

While there is much focus in the current environment around social media platforms and both the privacy and ethical issues involved with storing and using user data, the conversation about ethical data collection is not new. The entire scope of human subject research and the challenges of protecting those subjects have been a focus for researchers, regulators, and privacy advocates for decades. Indeed, more than 40 years ago the issue of ethical treatment for research subject made headlines with the exposure of the Tuskegee Syphilis Study^{xxiv} and the discussions that surrounded the Tearoom Trade Study.^{xxv} The former dealt with medical experiments done on African Americans at Tuskegee University starting in the 1930s, and the latter was a sociological study of homosexual encounters in public bathrooms in the 1960s. Both have the dubious distinction of being primary drivers of changes in rules for ethics in research. The outcry over the Tuskegee study led to the creation of the Office for Human Research Protections within the Department of Health and Human Services as well as federal laws and regulations requiring the creation of Institutional Review Boards for the protection of human subjects of studies.

Do such concerns apply to internal transactional or business data? Most of the research issues focused on the collection of data from subjects. While most businesses don't have "subjects," they do have data about people such as employees, customers, contractors, and business partners. What are the ethical issues that arise when collecting those data? It may appear that data on employees would be simple: businesses need to have information on their workers to execute payroll, provide insurance, and administer benefits and other administrative tasks for employees. Few ethical issues arise if the data that are collected directly support those activities. Some organizations may want to go beyond that, however, and collect additional information in order to do analytics to improve productivity and employee engagement. But are employees informed about that? Do they approve of such use? In areas where data are directly collected through a business process, organizations have an opportunity to outline why the data are needed or how they might be used. For data collected through scanners, cameras, sensors, and other devices, no such opportunity exists.

A strong data governance program can help address some of these concerns. While something as rigorous as an institutional review board will likely be unnecessary, an effective data governance program can provide guardrails for analytics projects as new data types and analytic uses arise. Regardless of which style of data governance framework an organization employs, a key component is the ability to communicate across business lines and the identification of appropriate roles and decision rights for the use of data. Can, or should, an organization use information from license plate scanners in a parking garage? Can, or should, an organization match their employee attendance records with medical appointment data? Can, or should, an organization match door scanner times for employee entry with employee time records? As more data privacy issues emerge, more privacy laws are discussed. The European General Data Protection Regulation^{xxvi} (GDPR) expressly grants rights to employees to know how their personal data are being used providing a legal basis on which to base ethical behavior. Even without such a law, conversations within a data governance context can help tease out the difference between what an organization can do and what it should do. It provides a forum for conversations about what senior leaders believe is appropriate use for data, what policies and organization might have or need for certain uses, and what risks might be introduced by undertaking a particular practice even if it does not violate laws, rules, or policy.

Ethics and Data Sharing

Once an organization has the data, they probably live in a particular department or area that needs those data for their business processes. Other parts of the organization may benefit from having those data but often there are no mechanisms for data sharing. For a lot of organizations, silos persist because there isn't a good understanding of how data activities in various areas or departments are related. There may be strong feelings of “tribal loyalties” where staff see themselves as part of a smaller office or departmental tribe rather than a member of a wider organization. Often ethical issues are based here: one department understands the data and the “right thing to do,” and there is no similar appreciation for that ability in other parts of the organization. This kind of microcosm thinking leads to, at best, a lack of understanding of how others interact with, use, or manage data. At worst, it provides a foundation for a lack

of trust across an institution as people in other tribes are seen as having incompatible goals or suspicious motives which may prove to be an impediment for data sharing. Even in organizations with a strong culture of ethics and integrity, the view of “us” versus “them” can prove to be a challenge for getting governance off the ground. Signals of this challenge can be found in concerns that others will use the data incorrectly, either through ignorance or malice, or the belief that one area is so special that others can't possibly understand or appreciate the data issues they face.

Regardless of the reason for it, lack of trust is a major impediment to a successful data governance program. Passive resistance from data stewards who are not convinced that data governance will provide the necessary safeguards can sideline or delay any activity that is seen to benefit the “enterprise” at the cost of the loss of local control (see Chapter 2). Creating a “code of data conduct” or “rules for responsible use” may help in both establishing baseline ethical uses as well as helping to build broader trust in the data community. Here are five C's that can help to define responsible use in a way that may help enable sharing (adapted from Mason, Loukides, & Patil, 2018):

1. **Consent:** Ask permission first, not forgiveness later. All data have access rules; even public data have an access rule which is “everyone can use them.” Data stewards need to make clear what the access rules and processes are for their data. If there are uses that are always prohibited, those should be made clear. If there are rules that apply sometimes but not others, outline general conditions. Data users need to be aware of and follow the access rules. Responsible use precludes asking forgiveness instead of permission, and expediency or immediacy of need is not an overriding consideration. If the data access rule does not address emergency access requests, a conversation with the appropriate data governance official and business leader is in order.
2. **Clarity:** Be clear as to the need for the data, what will be done with the data, who will be using the data, and for how long. Consent needs to be informed. Data stewards have the right to information about the use of their data consistent with security and privacy concerns. Users should expect to provide information about who needs the data, what they will do with the data, and how long to expect the activity to take. The level of detail the user will need to provide is likely to vary by data source. Stewards need to have reasonable expectations around the reason for

access: to break down data silos, suggest that the working definition for “need to know” should lean more toward “because there's no reason not to share” rather than “only share if you absolutely must.” If data are an enterprise asset, then they can be made broadly available subject to appropriate security measures.

3. **Consistency (and trust):** Users should do what they say they are going to do, and nothing else, so they can be trusted in the future. Once access has been granted to specific individuals for a specific purpose and specific length of time, the data user is expected to abide by the agreement. If one of the people working on the project changes, that needs to be approved. If a user wants to do something other than what was originally stated, that needs to be approved. In order to build trust around data, users need to be cognizant that the data steward is placing faith in their word.
4. **Control (and transparency):** Abide by the rules that the data trustee has set, and make sure they have a way of seeing that the rules are being followed. Data users need to be aware that access does not mean free rein. If there are specific rules and restrictions on use, they are in place for a reason and that reason should be respected. Where possible, data users should be transparent about what they are doing so that data stewards can feel comfortable relinquishing some of the control over the data that they manage.
5. **Consequences:** Those that don't follow the rules must pay the price. No one likes to talk about enforcement or punishment, but trust can't be established if there is no recourse for not being responsible. Ideally, data access policies will clearly state the consequences of a rule violation, but in instances where that is not the case, the least the user can expect is to have their access revoked immediately. Depending on the circumstance, legal or criminal action could result.

All of this sounds like scary totalitarian governance but it doesn't have to be. For data that are extremely sensitive, there will be much tighter controls and more detailed access and use policies. Most data will have very few restrictions on usage: many data stewards just want to know who is using their data and what they are doing with it.

Ethics and Analytics

Part of the trust issue around data stems from concern about people using data “incorrectly.” Even with a good understanding of, and policies for, ethical data collection and a foundation for responsible use as described above, apprehension may still exist for some data uses. For more than a century, people have noted the challenges with “lies, damn lies, and statistics” and today's environment is no different. Recall that the basic level of data literacy – the data reader – is suggested as a straightforward way to help users untangle the signal from the noise when understanding data. Also recall that data communicators are those able to tell stories with data. The unspoken assumption in that level of literacy is that the stories they tell are true and without distortion. The ethical implications of such an assumption must be explicitly addressed.

As we have discussed previously, standards and transparency about their use may often be put aside in favor of a definition, term, or calculation that paints a person, department, or organization in a more favorable light. No one wants to look bad and often data and analytics will highlight issues or challenges within an organization that some would prefer to keep hidden. Tales of strangely defined populations or quirky calculations that paint a particular picture are easily found. In fact, there are newspaper columns and podcasts devoted to helping elucidate challenges with data claims in the news. Corporations and other organizations can find themselves in a similar position, and staff and leadership need to think carefully about the blurry line between flattering and fraudulent. We have no hard-and-fast rule for how to stay on the right side of that line but instead restate the aphorisms to show how they might help highlight challenges when they arise:

- Just because you can, doesn't mean you should: can be helpful during discussions of what *can* be done to “fix” an unflattering representation that data might provide. It can be useful to remember that painting a prettier picture doesn't change the reality of the situation.
- If you don't want people to find out, you might not want to do it: If a discussion devolves into how to obscure a particular derivation or definition, this may be a helpful reminder of how challenging it will be to explain things if the truth is revealed.

Conclusion

In today's world, we are awash in data, enthralled by technology, and enticed by the prospect of what the combination of the two can provide. Data quality metrics can help organizations understand which data are best for what purpose and what the limitations for the use of those data might be. Data literacy is needed in order to make sure that organizations are well positioned to take advantage of the data deluge regardless of the technology train. Remember, data are an issue separate from technology, and data literacy is helpful even for those who will never need to draw a chart or write code. And last but certainly not least, data ethics are needed to ensure that data are used appropriately. Often the thrill of the possible clouds the judgment about the importance of the permissible. As data and the use of data continue to gain strategic importance for organizations of all types, data governance can help provide the environment to determine what can and should be done with what data and by whom.

Questions to Consider

1. What data quality metrics are discussed in your organization? Which are the most important? To whom and why?
2. What would a data quality program look like for your organization? What would it focus on? How would it work?
3. What level of data literacy exists in your department? What about organization wide? What level would be optimal?
4. What type of “data citizen” are you? What types are in your department? What types are needed?
5. Does your department or organization provide the tools needed for data literacy to succeed?
6. What data policies does your organization have? Do they address legal or privacy risks?
7. What is the culture around ethical data use? Do data users explain metrics and visualizations? Do senior leaders understand ethical issues?

Data Governance Maturity Models

Introduction

Data governance leaders and advocates are repeatedly challenged to prove that the investment in data governance initiatives is delivering measurable results. This is not as easy to do as it may be in projects with clear start and end dates and measurable goals. Without this information, ongoing and sustainable investments in data quality, integrity, and governance will become an ongoing challenge as senior leaders will steadily withdraw support.

Every organization, regardless of the industry in which it operates or the nature of its business, finds that it must answer the following questions. In fact, finding answers to the following questions is critical to survival:

1. What are our goals and objectives?
2. Can we measure and monitor our progress toward our goals and objectives?
3. What prioritized and methodical steps should we take to ensure that we are on track?

When it comes to answering these questions in the realm of data governance and management, we refer to it as an assessment of an organization's data maturity. Models that support finding answers to the above questions are referred to as Data Maturity Models. Maturity, in the context of data governance, refers to current organizational governance processes targeted at achieving the strategic and tactical data governance goals and objectives. Processes evolve and so does the data governance maturity of an organization. The quality, rigor, and

efficiency of processes continuously evolve from ad hoc, to formal, to monitoring established metrics, to full optimization.

Any assessment model must address all the elements that constitute the ecosystem of data governance. Periodically assessing how the overall organization is leveraging its data to achieve its mission and goals in an efficient and effective manner. It can provide extraordinary and invaluable insights into the overall health and well-being of the organization and its units.

Assessment begins by studying the data management and governance metrics for each unit. It takes a holistic view by compiling the net results of each unit into a comprehensive whole. This approach sheds light on the weak links in an organization and provides incentives to address these challenges systematically. The assessment of data governance is not different from many other organizational assessment methods. The idea is to objectively gauge an organization's current performance and measure the gap between the current state and desired state, minus the blame game. As in any other assessment effort, this broadly entails the following generic steps:

- Identify what data governance variable, processes, and systems to measure;
- Establish the link between what variable to measure and its impact on key performance indicators (KPIs);
- Identify how best to measure these variables and the strengths and weaknesses inherent in such metrics;
- Identify a broad range of stakeholders who are most knowledgeable about the issues related to this variable and its metric;
- Seek input (in multiple formats such as surveys and interviews) from a diverse, yet knowledgeable, group of stakeholders about the most appropriate variables and their metrics that truly reflect the organization's data governance health and well-being;
- Aggregate the variables that were measured and develop a strategic understanding of the issues and opportunities available for strengthening data governance in the organization;
- Identify the gaps between the current state of affairs and desired goal

state;

- Develop a roadmap to bridge the gap and methods and processes to continuously monitor the roadmap as business needs and environmental variables change.

The Three I's of Data Assessment

Data assessment can be a costly and time-consuming endeavor. It demands time, focus, and thoughtful engagement from a wide variety of internal and external stakeholders. It is essential for an organization to have clarity around the end game of an assessment initiative. What do we hope to achieve at the end of this assessment? Are we looking for evidence that a specific problem or concern exists or are we looking for insights to help us move the needle forward? Do we have the resources and commitment to address any weaknesses that the assessment reveals or will this become another interesting exercise? This is critical if the assessment is to be meaningful. In the absence of a plan forward, stakeholders will provide only limited feedback or partial truths. They will not provide insights that add value to the assessment.

We suggest there are three drivers, referred to hereafter as the three I's, that help organizations to identify the purpose behind an assessment initiative. These are:

1. Integration;
2. Integrity; and
3. Insights.

Data Integration

Data Governance is inherently a data integration challenge. Data that exists in only one department or group is only generating value within that part of your organization. The moment that data is shared across business groups, it becomes an enterprise asset that must be governed in order to protect it and maximize its value to the organization as well as enable

the organization to reduce costs in maintaining it. (Data Governance Assessment Discover the Business Value behind Data Governance)

If a key benefit of data governance for an organization is data integration, a highly dynamic, and frequently, messy endeavor, then the assessment should focus on the input and output gates of data, and the complicated chain of interunit data storage, retrieval, analysis, sharing, and modification processes that impacts the integration scenario at an organization. Data flows in and out of an organization are not always disciplined, structured, or predictable. When it is difficult to share data between applications and several operation systems, when data silos become the norm, and when valuable and insightful data that can tell a rich story about KPI and strategic issues remain inaccessible to decision-makers, it is worthwhile to assess the integration challenge through data maturity assessment models. The integration challenge is an enormous one that influences data quality and integrity, which in turn influences the quality of decisions. This is one of the great benefits of maturity models – it sheds light on the weak links in data integration, which directly and indirectly can impact the quality of organizational decision-making.

Data Integrity

Data integrity is not a one-time problem. Instead, it is a perpetual problem for many, if not all, organizations. If data are dynamic, data integrity issues will arise. This is inevitable. No organization can claim that 100% of its data has 100% integrity 100% of the time. Such an endeavor will be unreasonably costly with minimal returns. Refer to the detailed discussion on data quality in Chapter 6. Hence, it is essential to accept that lack of data integrity is a challenge that organizations must live with. However, what is essential is a critical understanding of the kind of data that must be free of integrity issues 100% of the time. Such data are often classified as critical data or metadata. Integrity errors in this category can have far-reaching consequences. The most serious threat is that it erodes trust in the quality and integrity of *all* data in the minds of decision-makers and not just the data that are corrupt. This loss of trust leads decision-makers to rely less and less on data and more and more on gut feeling and intuition. In such cases, an organization will be unable to reap the investments of time and money that it has made in collecting, storing, and

compiling data. In addition, the quality of decisions is compromised when stakeholders paint all data as unreliable and find shortcuts to arrive at a meaningful decision. If stakeholders in an organization frequently complain about a lack of trust in their data, an assessment that focuses on data is overdue.

Data Insights

Data insights as a driver of assessment can be meaningful for organizations that have achieved a higher level of sophistication in their ability to leverage data for competitive advantage. In such organizations, data integration and data integrity are not showstoppers. Decision-makers trust their data. Systems, processes, policies, and procedures are in place to support a comprehensive, detailed, and holistic view of the data that reside in the organization. Stakeholders can analyze data and deliver impact on key organizational metrics. Data are treated as a valuable resource, a currency that strengthens the market position of the organization. In such cases, the organization is focused on gaining deeper and richer insights from its data. What else can be done with data to strengthen market position? What additional insights can be gained through multiple richer perspectives of existing data? What additional data would help decision-makers enhance the quality and timeliness of their decisions? In short, the organization already has benefitted from its progressive agenda on data as a resource and now wants to leverage these benefits to the next level. The goal of data maturity assessment in such cases is (1) not as much to fix problems as much as it is to avoid future problems and (2) to strengthen its use of data as a competitive weapon. Conducting data assessment to strengthen the quality, timing, and quantity of insights has a direct bearing on the bottom line. Hence, organizations that understand the link between profits and insights are prone to invest heavily in ongoing data assessments and then following it up with strategic actions and initiatives to reap greater insights from existing and new data.

In summary, the triggers for data assessment can be data integration, data integrity, or data insights. However, these are three interrelated and overlapping concepts. Data integrity and integration can often lead to richer insights. Data insights can lead to new ways to integrate and analyze data. Hence, the impact of pursuing a given purpose certainly can have a positive impact on the other two. We list these three triggers only as primary drivers of assessment, not as mutually exclusive or as a comprehensive list of assessment drivers.

What Is the Purpose Behind Assessment?

It may be tempting to engage in a data assessment model because everyone else is doing it. Or, it may be one of those activities that an organization engages in at regular time intervals, or one that is triggered by top management. Regardless of who initiates the data assessment initiative in an organization, clarity about the goals and objectives of the initiative are key to success.

Descriptive

The purpose, in this case, is to describe clearly the current status of data as an asset. This approach works for organizations that have a clear set of criteria to measure their data maturity. The goal is to take a snapshot in time and describe the results to top management and other stakeholders. For organizations that have invested in developing clear and measurable criteria that have been clearly articulated and communicated to key stewards and stakeholders, a maturity assessment to describe current status is meaningful. It is essentially a report card against a set of pre established metrics.

Prescriptive

Any meaningful maturity assessment should provide guidelines for improvement and enhancements. The purpose of any assessment exercise is to get better at what one is doing! However, sometimes the purpose behind assessment may be primarily to seek innovative ideas from key stakeholders and to methodically address weak links or hidden problems. Stewards and stakeholders may use the assessment exercise to provide deeper and richer insights into how the organization can better leverage the data it has. This pursuit is more appropriate for organizations that already have a strong record of using data to achieve its strategic goals.

Predictive

Sometimes organizations know that trouble is brewing. Data are misused or abused. Data are unreliable and decision-makers rely on gut feeling and

intuition. The quality of data is poor and there is a lot of finger-pointing. Regulatory and compliance issues are starting to surface. In this case, the best way to get everyone's attention is likely through a data assessment exercise that outlines the economic, political, and social consequences of ignoring the above mentioned issues. On the other hand, predictive data assessment can also be used to inspire greater investments in data governance. Data governance advocates can show how continuing on a positive slope can help the organization achieve its strategic goals and capture market share.

Comparative

The goal here is to analyze and compare the ranking of an organization compared to other companies in its industry. It can also be useful to compare one internal unit with another or one subsidiary with another. Such internal or external benchmarking can be highly useful. However, it is not easy to do. Rarely are data about the competition available along the same metrics or parameters to facilitate a meaningful comparison. Still, where and when possible, a comparative assessment can be an eye-opener.

In summary, the goal is not to box oneself into one category or the other. The goal is to adopt a meaningful approach to the goals and challenges that the organization faces; is to actively and purposefully mix and match; tweak and refine and re-refine until the assessment exercise is exciting and meaningful to key stakeholders, if not all stakeholders. The goal is to achieve a balance between the complexity and strengths and weaknesses of a given model versus the complexity, goals, issues, and challenges that the organization faces. Investing time in selecting the right model and customizing it can reap extraordinary rewards and make assessment a central theme in all strategic discussions and investment proposals.^{xxvii}

There may be other compelling reasons to engage in data maturity assessment.

- Future investments in data governance rely on compelling evidence that existing initiatives are impactful and deliver results. A data maturity model can provide a comprehensive assessment of the data health of the organization, along with individual units.
- Existing data governance initiatives must be measured for their efficiency

and effectiveness. For example, if new data governance policies and procedures are being incorporated into the project management life cycle, studying its impact through data maturity assessment can be useful.

- One of the greatest benefits of a data maturity model is that it sheds light on the weak links in the system. It gives a diverse set of stakeholders the opportunity to share their experiences, ideas, and concerns. It turns the light on how an organization uses one of its most valuable currencies.
- The best policies, procedures, and processes can become dated. What once was useful and meaningful can become cumbersome or even unhealthy. Assessment can lead to a health check of these elements.
- The results of data maturity assessment initiatives can provide input to training curriculum and programs. Repeated breakdown in what employees do with data or don't do with data can help trainers address the gaps.
- Finally, one of the most notable contributions of assessment is that it ensures that the organization is meeting its regulatory and compliance requirements. By addressing the weak links before it becomes a breach, organizations can avoid severe financial and political consequences of regulatory violations.

A Word of Caution

Before we present different maturity models, a word of caution is in order. Maturity models are just that – models. Not every model will work for every organization. Even if there is a model that meets the needs, generally speaking, it is understood that some level of tweaking and customization is essential to reap the full benefits of what the model has to offer. The questions embedded in each model are designed to derive insights and identify weak links. Often, in our work, we find that the questions may have to be reworded to align with corporate language and culture. A canned questionnaire almost never works and in fact, the results of such a survey can be misleading.

Organizations need to take the time to find a model that works and then to invest the time and resources necessary to customize the model to meet their current needs and priorities. Organizational structures, stakeholder expectations,

past successes and failures, regulatory and compliance requirements, the scope of the assessment (enterprise, business unit, project, process, people, etc.) can shape the quality and purpose of an assessment. While the technical and managerial aspects of data governance are essential, it is equally important to bring an understanding of the cultural norms and practices of the organization to the model. Finally, even a customized model has a shelf life. Hence, we recommend reviewing the model once every few years to ensure that it is appropriate and timely to the shifting data needs and challenges an organization faces.

Once an organization has settled on a model and an approach, it is time to have a very strategic conversation: what is the optimal level of maturity? Such a question might seem nonsensical: of course the answer is to get straight to the top! Everyone wants to get the highest grade or top mark. Why wouldn't an organization want to get to the highest level of maturity? The short answer is that it may not pay to do so. Below we describe different models that have different scales and levels of maturity that an organization may achieve. There will be tradeoffs, costs, and benefits, to moving from one level to the next. Like all business decisions, evaluating what kind of return is needed on an investment to make it worthwhile applies to data governance maturity decisions as well. As we mentioned previously, models are abstractions with a shelf life and striving to get to an unachievable or unhelpful level of a particular model without a strategic evaluation of that goal may not yield the best investment for the organization.

Five Levels of Maturity

Most data maturity models include five generic levels. There is no logic or compelling argument as to why we have five levels and not six or ten. Hence, the five levels should be considered as just a guideline that attempts to track the predictable growth in steps or evolution in stages in the area that is being assessed. This is the reason why maturity models are sometimes referred to as stage models. A few caveats are appropriate before we dive into the popular five levels:

- As stated earlier, most data governance and data maturity models present five to six stages. Remember, this is not the only valid scale. Any organization may decide to expand this list.

- The goal of implementing a data maturity model is to assess the data maturity of the entire organization at a given snapshot in time. This is equivalent to assessing the health of a given population segment. It gives a rich perspective of the overall landscape. It can help decision-makers step back and study how the landscape has changed since the last assessment. A numerical grade of three or four at the end of an assessment can reveal roughly where the organization falls on its climb toward becoming a data-driven organization. Nothing more, nothing less.
- It is difficult, even questionable, to identify levels in a model where there is truly no end state – there is only a journey of progress toward a dynamic goal, an ideal state that is frankly, unachievable. Hence, it is more meaningful to study the results of a maturity model through the lens of change and growth.

We discuss these levels in this section.

1. **Initial:** The organization is new to data governance. It has limited to no knowledge about the best practices in its industry. It performs only limited activities that ensure the quality and integrity of the data. Data are not treated as a precious resource. Senior management is not a strong advocate for data governance because of their own limited knowledge and understanding of this domain. There is no data strategy and stakeholders are not fully on board.
2. **Repeatable:** The organization is getting its feet wet in data governance. There are pockets of data governance activity in different units throughout the organization. These units are starting to establish policies, procedures, and processes so that data governance activities become repeatable and are considered the norm. These units are starting to reap the benefits of its data governance activities.
3. **Defined:** The organization has a documented standard for performing the assessed activity or activities consistently and using applicable tools effectively. Employees understand that data governance is not a passing fad. There are many well-established policies and procedures, and many units use some, if not all, tools of data governance.
4. **Managed:** The process for data governance is established, tracked, and

monitored. Employees understand their role in data governance and know that they will be held accountable. Employees are highly educated and informed about policies and procedures, which are treated as mandatory, not optional. Data governance goals are established for each unit. The recommended tools are in place and being used consistently across the organization. Standardization is the norm.

5. **Optimized:** The activity of data governance is continually reassessed and improved. Any loopholes or cracks are not sidelined or brushed aside. Instead, these are addressed openly and immediately. Top management is not interested in hearing only good news but is open to learning from failures. Maturity is tracked and measured across all units in the organization, and unit leaders are held responsible for leveraging data to achieve their KPIs.

Data Maturity Models

In this section, we present a few popular data maturity models. It is our intent to describe only a few widely referenced maturity models. We provide a brief description of each model and a source where one can find more information about the model. We also provide several resources for a comparative analysis of the models described in this section. There is a large volume of public information on maturity models. It is the responsibility of the assessment team to evaluate the models, overlay it with unique factors and cultural norms that apply to a given organization, customize the model for intended use and purpose, and once this is done, start the process all over again when the time is ripe for the next round of data assessment. Why? Because organizations are dynamic living organisms, and hence a model that may have worked well a few years ago might still need some refinement to reflect the new reality.

The Mother of All Maturity Models: The Capability Maturity Model

The Software Engineering Institute launched the Capability Maturity Model (CMM) in the early 1990s. Since then, we have had a proliferation of maturity models in different fields. Carnegie Mellon University's original contribution to

maturity models in software development has become the gold standard for areas other than its original intent. The CMM was developed to assess the gap between the current state of affairs in terms of policies, procedures, and processes, and business activities in software development by government vendors and contractors versus the desired goal in terms of building an efficient and error-free environment for software development. However, the flexibility and power of the model lie in its adaptability to environments other than just software development. For example, today we have maturity models in knowledge management, information technology, process management, and business intelligence, to name a few. Many software companies and consulting companies have also developed proprietary maturity models that provide insights into how to advance the use of data, governance, and related processes in an organization. Hence, it is frequently discussed in the context of assessing the maturity levels of data governance in an organization. Regardless of the domain, the goal is to take stock of the current situation, identify opportunities for growth and progress, and develop a path forward.

The Data Management Maturity Model

The Data Management Maturity (DMM) focuses on six program areas:

1. Data management strategy: including communications, business cases, and program funding.
2. Data governance: including business glossary and metadata management.
3. Data quality: including strategy, data profiling, assessment, and cleansing.
4. Data operations: including requirements and lifecycle management.
5. Platform and architecture: including architectural standards, data management infrastructure, and data integrations.
6. Supporting processes: including measurement and analysis as well as process, risk, and configuration management.

EDM Council's Data Capability Assessment Model

The Data Management Capability Model (DCAM) is a data maturity model that was developed by the Enterprise Data Management Council. The word “Capability” in the model refers to the emphasis that the model places on the capabilities necessary in order to do more than just an assessment: it focuses on a portfolio of capabilities that an organization will need in order to be able to move up the data maturity ladder. In particular, it addresses the business strategy, organizational structures, technology, and operational best practices that must be in place in order for data to become a valuable asset that is successfully leveraged to achieve the goals of the organization.

The DCAM is divided into eight core components:

1. **The Data Management Strategy:** As we have stated earlier, without a clear data strategy, organizations will flounder. We recognize that data strategy is not quick or easy to develop. However, this model emphasizes the importance of strategy and discusses the essentials of a good data strategy.
2. **The Data Management Business Case and Funding Model:** Many models skip this important and fundamental phase as to how data governance advocates should present a strong business case and the elements of a good funding model. This phase also emphasizes how critical it is to engage and win the support of senior executives and key stakeholders prior to embarking on an assessment exercise.
3. **The Data Management Program:** The focus here is what is necessary for the organization to establish a sustainable Data Management Program. This element of the model focuses on the critical question, “Is every operational part of the organization geared toward treating data as a valuable asset that has a direct impact on the success and sustainability of the organization?” In other words, how do we embed core data management principles in every operational function within the organization?

The function of data management is derived from an understanding of business objectives and organizational priorities as well as knowledge of how

data flows from initiation through validation through enrichment through transformation and into consuming applications. In many environments, data content is understood as a manufactured product and flows through the organization as part of a linked process. It is this linked nature of the process that highlights the collaborative components of data management. The goal is to instill a sense of collective ownership of data quality among all relevant stakeholders.

This phase of the maturity model examines if all stakeholders understand, value, and implement the core principles of data governance.

4. **Data Governance:** A data maturity model ensures that there is effective data governance. This phase of the DCAM focuses on the quality and rigor of policies, procedures, standards, and processes that are in place to deliver the vision and mission of data governance initiatives and strategies. It ensures uniformity and consistency of data governance principles across diverse stakeholders.
5. **Data Architecture:** This phase of the assessment focuses on preserving and protecting the consistency and appropriateness of data usage, its quality, and integrity. Data exist to satisfy business requirements, and data architecture is the foundational element to link data with requirements. This phase looks at how data, metadata, and critical data are defined in logical domains and physical domains. It ensures that the definitions of metadata (descriptive, structured, and administrative) are free from inconsistencies and that critical data are clearly defined and communicated to all stakeholders. This phase is therefore critical as it examines a wide range of issues, including data duplication, clarity of lineage, shared data dictionary, and data ownership.
6. **Technology Architecture:** This phase of the assessment takes its cues from the data architecture assessment covered in the previous step. It deals with the physical IT infrastructure that is essential to support a wide range of internal and external decisions. Technology architecture

assessment in this model focuses on efficient storage, easy access, optimal processing speed, and appropriate security and flexibility and adaptability to ongoing technological advances.

7. **Data Quality:** Most readers by now are familiar with the critical importance of data quality and its centrality to effective decision-making. This phase of the assessment focuses on the key question, “does the quality of the data fit the purpose for which it is used?” This is frequently referred to as “fit-for-purpose” and while it may sound simple and elegant, implementing this idea throughout an organization is a challenge. Quality is closely linked with the control that is exercised throughout the data supply chain from initiation to disposal and the processes behind it that drives quality.
8. **Data Operations:** This is the phase that pulls it all together and sets the tone and efficacy of the core principle that data are an invaluable asset that must be managed and protected. This principle must be preserved and protected throughout the data life-cycle process and must be “integrated into the overall organizational ecosystem.” This model advocates the assessment of data management capabilities, the strength of stewardship embedded throughout the Data Management Life Cycle, and “the integration of data management into the information ecosystem.”

(Source: https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf)

IBM Maturity Model

The IBM Maturity Model measures the data health of an organization along six dimensions and five levels of ranking. In our opinion, it is one of the few models that does not shy away from the abstract, or that which is difficult to define and measure. It recognizes the messy nature of data assessment and organizational dynamics.

1. **Strategy:** This dimension of the assessment model explores and evaluates if data strategy is an ad hoc standalone strategy or if it is integrated and embedded in the core business strategy. At the highest level, organizations use data to drive innovation and to assume a

leadership position in the marketplace. At the lowest level, data are simply a resource to report financial data and regulatory mandates.

2. **Information:** This assessment phase examines the role and contribution of information to organizational success and market power. Across the five evaluation rankings of the model, an organization should assess if the information is being used to observe, manage, improve, differentiate, or at the highest level, as a strategic asset.
3. **Analytics:** The role of analytics takes center stage in this assessment model. Rather than merely advocating the use and benefits of analytics, this model incorporates it as an important and value-added element in data maturity assessment. It explores the level of usage and the strategic nature of analytics in decision-making. Is analytics being used at the lowest level to simply describe what has already occurred? Is it being used at the next level to describe why something occurred? Is it being used to predict a future event in some business activity? Or, is it used to maximize business value or at the highest level to achieve a competitive advantage by automating critical business processes for strategic advantage? Clearly, the ranking along this dimension is likely to vary from one business unit to another. It is difficult, if not unrealistic, to give a single score to a multifaceted organization. Still the power of asking how a unit is using analytics to achieve its goals can be meaningful. It can also lead to more proactive usage of analytics.
4. **Culture and Execution:** This is probably the only data maturity model that has culture and execution as a central component of its assessment. As we have stated before, the success of data governance is closely intertwined with the culture of an organization. In this phase, does the assessment focus on the link between organizational culture and data-driven decision-making? Is the organization ruled by the whim and preferences of an individual (the lowest level) in which case the organization fails to derive the benefits of effective data governance? At the highest level, key stakeholders are dedicated users of analytics and data-driven decision-making with the singular goal of maximizing business value. The scale shows how culture can influence the ability and intent to use analytics for competitive advantage.

5. **Architecture:** This refers to the data architecture and technology architecture concepts that were discussed in the previous chapter and hence, will not be duplicated here.
6. **Governance:** This dimension focuses on the extent to which data governance policies, procedures, processes, and systems have become embedded in the organization. From arbitrary and ad hoc policies, to clearly articulated and communicated rules and regulations, systems and processes that are grounded in a strategic, systematic, and structured approach to governing the entire information ecosystem, this assessment tool allows an organization to not only assess its current status but to also follow the trajectory of progress that it can make.^{xxviii}

Conclusion

There are several great resources that present a comparative analysis of different maturity models. We have listed two resources below. As we stated earlier, no one model can do justice to any organization. Each model has its unique strengths and weaknesses. It is difficult to explain an elephant in three words or less. Similarly, it is difficult to summarize and assess the several hundred elements that constitute the crux of data governance. Data governance is a messy business that functions within a dynamic business world with its ever-changing needs and challenges. Hence, it is prudent for those who embark on the assessment exercise to first and foremost study and review the maturity models. An impartial and objective view of the pros and cons of each model will assist in customizing the model to meet the unique needs of an organization.

Questions to Consider

1. What drivers do you see in your organization for doing a DMM assessment?
2. Which model do you think would work best for your organization? Why?
3. What level of maturity do you think your organization is at now? What level do you think you should aim for? Why?

8

Summary Case Studies

Case Study #1

Introduction

Our first use case is a financial services firm that is a leader in the US retirement industry with highly innovative products. The organizational structure is very typical for the industry: CEO, President/COO, CFO, executives and VP under each business vertical. There are two locations, with one primary home office and a satellite office in a nearby city. The primary business verticals are finance, investments, and Six Sigma under the CFO. Actuary, IT, PMO, and Operations under the COO. Sales, marketing, and distribution under the Head of Distribution. Legal and Compliance under the Chief Counsel. And just in the last few months, they have added a Chief Digital Officer, and the new head of Data Governance and Quality reports to this position.

Their vision is to be a leader by providing solutions across a broad set of retirement markets and wealth segments. Data Governance allows the business to make fact-driven decisions around new products to create and sell, as well as provide peace of mind to existing product owners that the data provided regarding their account are accurate and secure.

Insurance and financial services are by nature fairly conservative entities. Decisions are fully vetted and discussed prior to implementing changes to ensure the path meets both short-term and long-term needs of the company.

Data Governance Need Identified

A recent spinoff from the organization resulted in the creation of a servicing company. As part of the organizational split, the servicing company owned and maintained the IT data and infrastructure for both companies. In 2016, it was decided that the technology and infrastructure needed to logically separate and each entity would have its own. Up until this point, the original financial services company had only a couple data repositories, and they were built for specific business units like sales.

The new information technology area of the existing company started to build its own central data repository for storing their data. In the process of setting up this data warehouse, IT was trying to engage the business units to give direction on business context of the data feeds they were receiving from the servicing company and other vendors for the purpose of properly transforming and storing the data.

It was difficult trying to get the business units, who have little capacity, to step through this arduous process, and IT found themselves making assumptions and decisions that really shouldn't belong to the technology teams. The VP of IT reached out to the head of the Six Sigma department to see how it could step in and help to establish data governance.

Process

First steps were really for the Six Sigma team to understand the fundamentals of data governance. The company has data retention policies that are set and upheld by legal and security policies that are upheld by the Chief Compliance Officer, but there was no structure around governance for the purpose of data insight. The team started focusing on that issue.

First, they mapped existing data flows and found multiple teams receiving the same data files but transforming the data in different ways. They found reports that did not reconcile between finance and sales. They really tried to understand what pain bad data or lack of data was creating for the teams.

Next, they sat down with IT to discuss best practices for the data flow. How could they eliminate extraneous data stores and old access databases? They needed a plan that they could communicate with the business units.

Once developed, they took these ideas on a “road show” with all the business units to hype the idea of sourcing data once and reusing it multiple times. The message was that users could have easier access to data and there could be data quality measures set by limiting the number of places data are stored. This

element was a key point in the discussion: most of their data are supplied by vendors, and they needed a way to measure the quality of the data they were receiving.

After the initial discussion on strategy and approach, they were ready to consider tools. The first need they identified was for a data glossary and data quality tool. They spoke to six different vendors and narrowed the search down to three they felt would best fit their needs. They had each vendor spin up a sandbox region and completed training and a three-week proof of concept on each tool. The Six Sigma tool, the Pugh Matrix, helped to provide quantitative measures of the fit for each tool. ¹ After the trial runs and tool selection, they went through the challenges of contract negotiations and implementation.

During the tool selection process, they were also working on the people aspect of data governance. They created a data governance working group consisting of the VP of Six Sigma, VP of Audit, VP of PMO, VP of IT, a Six Sigma black belt, and several data super users, including the head of BI. The group met on a weekly basis to discuss data policies, projects that needed to be completed around data, and any problem issues that came up. The group also worked with all business unit leaders and identified the data stewards in each area. After the identification, the group began having monthly meetings with the stewards to discuss the role and responsibilities of being data stewards.

Current Data Governance Landscape

The groundwork done by the Six Sigma group and the data governance working group resulted in data governance structure comprised of the Chief Digital Officer, Head of Data Governance, Data Analyst, and a team of data stewards from each business unit that have other full-time jobs. The Six Sigma Black Belt stepped into the new role to head data governance. While the project was initially championed by the CFO, it is now really driven by the Chief Digital Officer.

The organization now has two different working groups. One group is at senior level, VPs and above, that is more strategic and meets twice a month. The other group is a more tactical working group that meets weekly. Two meetings per month are used to go through data definitions and quality rules using the data tool they selected. The other two weekly meetings are used to discuss issues or project work around data.

The current approach has been to start working on data that contain personally identifiable information (PII) as forthcoming legislation has made this a priority. They concentrate on defining business terms using a seven-step process:

1. Write the business definition.
2. Write the business rule for the term.
3. Consume the metadata for the term from any repository that holds that term.
4. Write a data rule from the business rule that can be applied to the actual column of data.
5. Measure baseline quality and set tolerance levels.
6. Send failing records for remediation.
7. Map the lineage of the data from ingestion to any outputs.

The results so far have been promising. After only 30 days of use, the tool has strong support. The stewards had goals added to their performance plans around creating definitions and actively participating in the meetings. The goal for the first year is to have all PII data complete, as well as 60 additional terms that the business considers critical data elements.

With enough experience over an extended time frame, they plan to work with vendors to have agreements in place to cover completeness and accuracy. The current data contracts only set limits around timeliness, and much more information is needed to adequately quantify data issues for the vendor discussions. In order to be successful, the organization will need to be able to provide quantitative measures of quality to prove the need for improved controls within the vendors' processes. Their data governance efforts will enable that and could possibly prove beneficial not just to their organization but to others who use data from the same vendors.

Lessons Learned

- Six Sigma is a great team to start building data governance. They work

with all of the business units from a neutral perspective, so there is no agenda in place to drive things toward a particular solution. They are there to listen to the needs of all business units, including IT, and build a sustainable model.

- Make sure you are aware of how data issues have created pain for each business unit. When trying to build support for data governance, be able to give them specific examples of how it will solve their personal pain points.
- It will take longer than you thought to get things moving and that is OK.
- This is not a PROJECT with an end date and that needs to be socialized from the very beginning. We are never “done” with data governance.
- Look for tools that have accelerators like auto-discovery and built-in industry glossaries.

General Advice

- Assess your current State:
 - Map where your data live
 - Who is using the data? Who is managing the data?
 - Do you have any governance?
 - Do you have any quality checks?
 - What about vendor data?
- Gain Support
 - Top-down communication means you need to equate data strategy with corporate-level goals and try to quantify the cost of bad data
 - Bottom-up communication is helpful to understand the pain points bad data create for each business unit and allows for discussions on how data strategy can address those pain points

- To really get people on board, it is important to give realistic expectations and timelines for improvement. Also, make sure everyone understands this is not a project with an end date.
- Execute on Strategy
 - Make the data strategy work part of existing projects
 - Add data strategy tasks to all project plans
 - Include funds to support the data work in all project estimates
 - Make data governance and strategy work a requirement for any moves to production
 - Remember: It's a marathon, not a sprint!

Case Study #2

Introduction

Our next use case is from the higher education sector. Rochester Institute of Technology (RIT) is a private university that was founded in 1829 in Rochester, New York. RIT has nearly 200,000 students, 75% of which are undergraduates, and about 4,000 faculty and staff. They have global campuses in China, Dubai, Croatia, and Kosovo, and are the third largest producer of undergraduate STEM degrees among all private universities in the nation. With the variety of activities, they have taking place on multiple campuses across the globe, the university recognized the importance of tackling their information management practices in a more formal way. Most institutions of higher education have a long history of analytics aimed at understanding their key populations of students and faculty members as well as assisting with decision-making in areas such as admissions, curriculum assessment, staffing, student activities, finances, and alumni relations. Such activities are generally located in a department for Institutional Research (IR), a key group that works with a variety of other departments on campus, especially university information technology staff. For many universities, data governance, or the recognition for its need, often begins in either the IR or IT areas. This was the case for RIT: when they undertook to

create an enterprise information management program, it would be spearheaded by the manager of their Enterprise Data Services in the Information Technology division as well as the Associate Director of Institutional Research, Data, and Analytics.

Like many higher education institutions, decision-making at RIT is decentralized. The culture places high value on creativity and innovation, not just on the academic/research side but across the institution. There's a lot of experimenting going on. Combine this with decentralized budgets and the consumerization of IT and it makes it difficult to standardize information collection and practices so information can effectively be shared.

Data Governance Need Identified

The first need for data governance was identified in 2013, and like many other organizations it was triggered by a specific problem: the proliferation of business intelligence tools and the risks associated with the existence of multiple platforms being deployed to undertake similar activities using the same data across many areas of the university. From the beginning, the program was a partnership between IR and IT. However, they found it difficult to make progress because the effort was inherently tool focused: how do we address the variety of tools that are being used. This question was seen as strictly an IT issue which made enterprise-wide progress a challenge. The business areas cared more about business outcomes and what data lived where. A key learning from this first attempt was to focus on the business outcomes.

In 2015, they launched a second attempt at enterprise information management. Learning a lesson from the first effort, this round was very clearly business focused. The discussions were around data governance not technology governance. They created a data definitions committee and a data integrity committee to bring standardization to their data and understand their data quality issues. They built data models to describe their information and ensure better understanding of all the information they had across the university. While this work definitely addressed the lessons learned from the first effort, the program still stalled. The data modeling work tried to capture every relationship between all the data elements across campus. In a large decentralized organization, this kind of approach is not only unwieldy but is rarely successful as it takes a great deal of time and effort to even get started. This second round of work taught the team that trying to model, document, and describe everything all at once takes a

lot of time and energy without the corresponding payoff. RIT realized that they needed a more strategic approach.

Once the need for a strategic rethink was recognized, senior leadership for the IR and IT departments held a full-day off-site strategic planning discussion in June 2016. They took a good look at where they were, where they came from, and what they wanted. They did an inventory of what they had and what they wanted and created a diagram that is now affectionately known as the “Big Mac.” Fig. 8 shows the parts of data governance they had in place, the colored items, and the pieces they still needed, the white items.

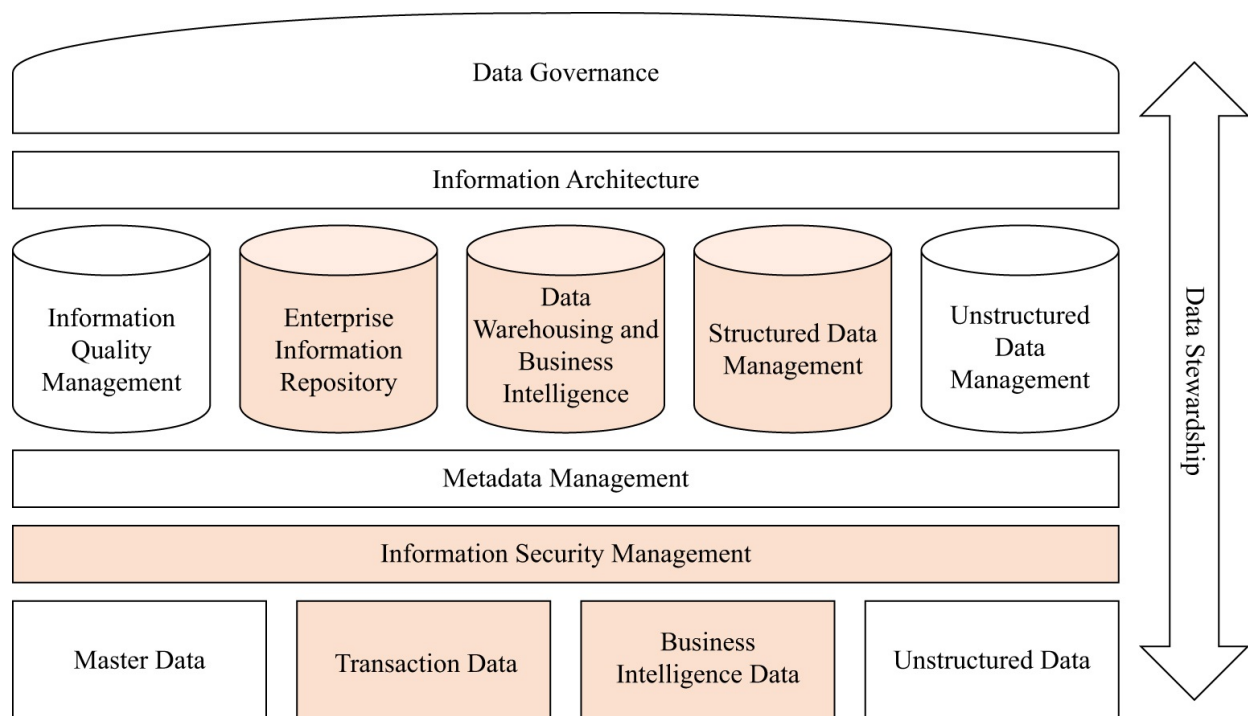


Fig. 8. Enterprise Information Management (EIM) Model at RIT.

In order to effectively plan to put in place the pieces that they realized they needed, they worked to create and hone vision and mission statements that were pithy, meaningful, and captured the right sentiment. The vision for RIT's Enterprise Information Management program is that “EIM provides an integrated data environment where information is delivered to those who need it, when they need it, and how they need it.” The supporting mission statement provides more detail on how to achieve the vision: “To advance EIM by establishing data governance, defining roles and responsibilities, promoting

awareness, and implementing technological solutions.”

Process

Now that they had their mission and vision clearly stated and aligned, it was time to work on how to actually work toward their fulfillment. First, they needed to understand what oversight framework would make sense for their organization. They wanted breadth and reach across the university while still retaining agility. Eventually they decided on the triangle shape for their governance format, as shown in [Fig. 9](#).

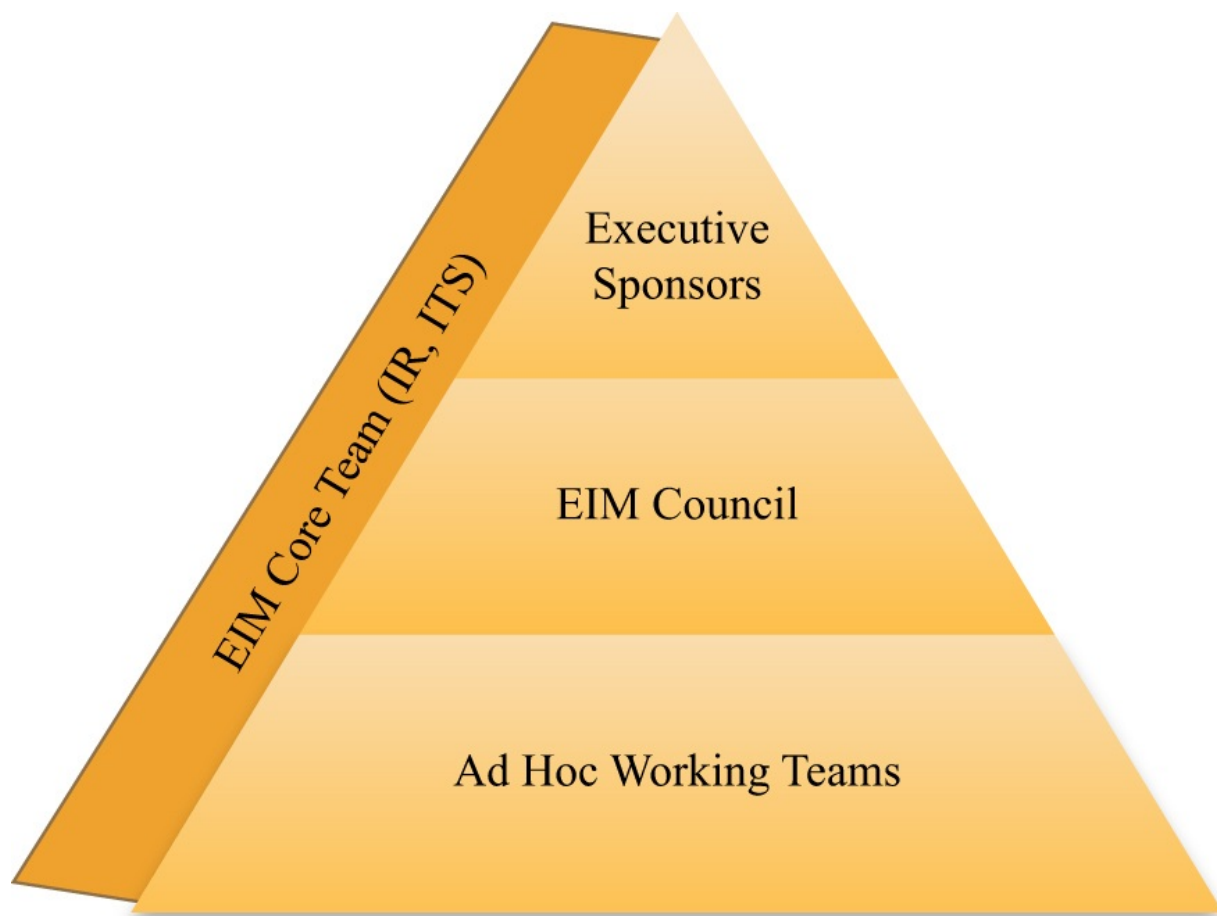


Fig. 9. EIM Organizational Structure.

They identified their executive sponsors as six VPs from across the University and then worked to set up the EIM Council: the tactical layer of their

structure. They targeted members at the Director level as staff who would be close enough to the work to feel the pain but connected enough to the leadership to do something about it. They selected representation from their key stakeholder domains: Enrollment Management, Registrar, Development, Student Affairs, Finance, Human Resources, Academic Affairs, Marketing, and the Information Security Officer. In total, the council has 15 people who meet once per quarter. They are regularly consulted and informed about data governance issues and are charged with staying connected to the Executive Sponsors as well as helping to identify staff for ad hoc working teams who comprise the operational level below the EIM council. Supporting the EIM work is the Core Team: a small group of IT and IR management who meet twice a month and are accountable for the program's success.

With the direction identified and the oversight framework in place, it was time to get started. One of the first decisions the new Council had to tackle was identifying which missing pieces of the Big Mac diagram to address first. They decided to take on data governance and data stewardship as these are really in the middle of all data management. Data governance was seen as key: it was difficult to see how other efforts could be successful without it. They decided it was the best way to bring critical people along for the ride and help build ownership among the relevant stakeholders. Therefore, a Data Governance Working Team was the first of the ad hoc work teams created as part of the operational layer of the framework.

Once formed, this team decided to focus on three key things: create a data stewardship organization; develop a data governance policy; and launch a sustainable data governance program. The first steps to achieving these goals centered around developing a data domain model and outlining data stewardship roles and responsibilities then creating a data stewardship organization and a data governance policy. For the data domain model, they started grouping our data by subject areas beginning with student data because they already had a robust set of student data in their data warehouse. Student data came in three flavors: prospective student, current student, and former student or alumni. After that, they started to identify other topical areas, working to define different types of data across the organizational silos such as IT, human resources, and financial data. [Fig. 10](#) shows the results of that exercise.

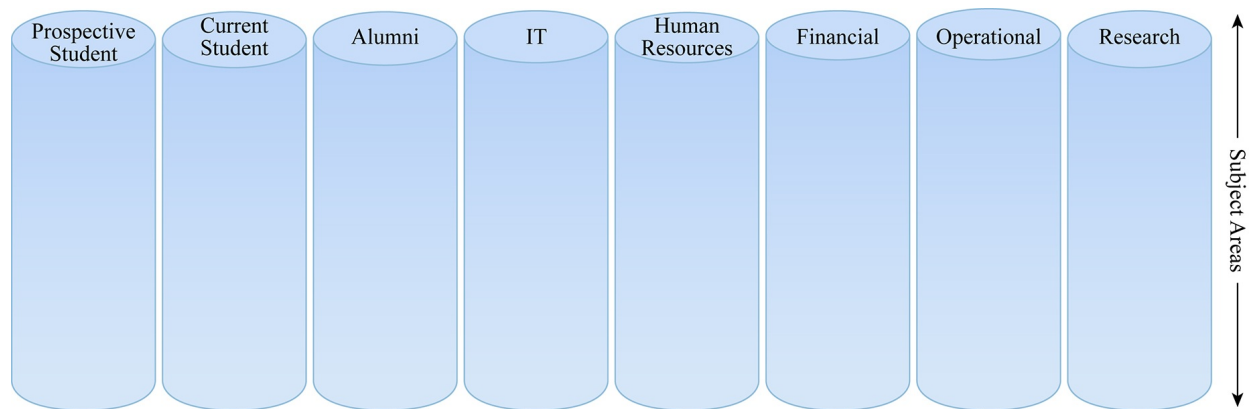


Fig. 10. Initial Domain Model.

Once these subject areas were identified, they started to think about the relationships between the domains and what that might mean. For example, Prospective, Current Student, and Alumni are all versions of “student,” and “Human Resources” and “Financial” both contain information touching the role of Employee. As a result of this introspection, they were able to identify two different levels of domains. A Level 1 domain is contained within a single subject area, whereas a Level 2 domain spans multiple subject areas. Fig. 11 shows the resulting Level 1 and 2 domains identified.

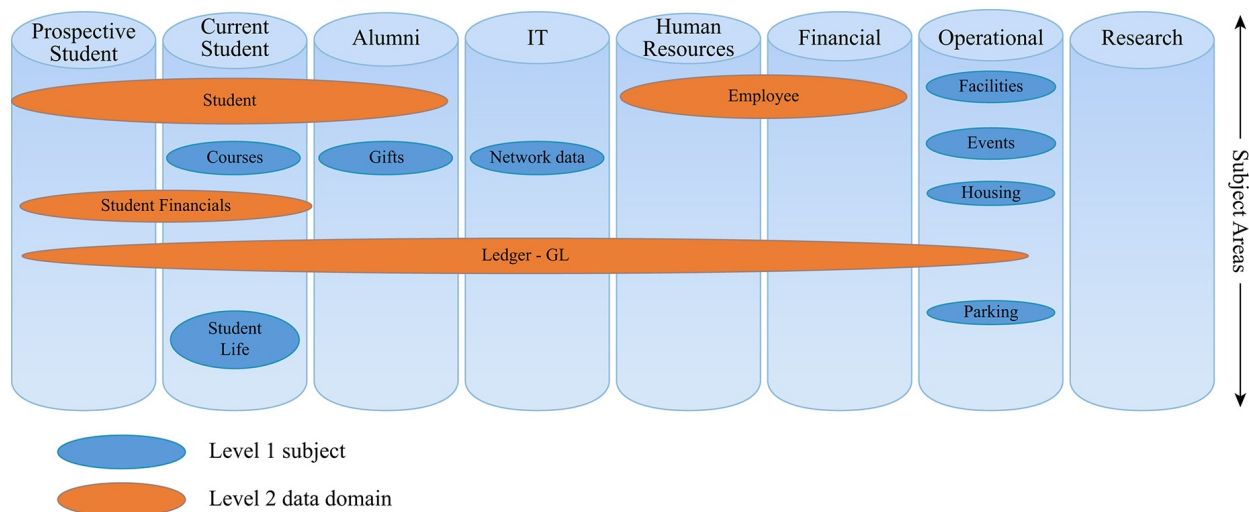


Fig. 11. Revised Domain Model.

With the domains defined, they turned their attention to stewardship roles. In deciding what roles they needed, they turned to their EIM governance model and

decided to define a role to correspond to each level. At the top level, were the Data Trustees: executive leaders with policy-level responsibility for managing a major area of the university's information resources. For the tactical level, they defined the role of Domain Data Stewards who are senior-level University officials with planning and policy-level responsibilities for data in their respective domain(s). These Stewards have oversight of data management for subject areas within a domain, and they identify the participants for the tactical-level role: the Subject Data Stewards. These stewards are the University staff who have responsibility for the capture, maintenance, access, and usage of data for specific subject areas. They have hands-on expertise with acquiring, valuing, creating, storing, using, deleting, and archiving data in their respective subject area.

Once the roles were defined, the people to fill those roles needed to be identified. They started with the middle layer, working to identify the Domain Data Stewards who were the same director-level staff as were on the EIM Council. Once those roles were filled, the EIM Core Team worked with them to reach out to the VPs and people at the Provost level to explain the Executive Sponsor role and ask for their support. With the two upper levels defined, the tactical level was next. The Domain Data Stewards recommended Subject Data Stewards for each subject area. Some Domain Stewards named themselves and others named other people within their functional area of the organization. Some roles are still being clarified. For example, Student Life is a Level 1 domain within the Current Student subject area. Does it get its own steward? Or do they only have the Student domain steward?

They then moved on to the final deliverable for the Data Governance Working team: defining a Data Governance Policy. The team met weekly over the course of several months to get this done. This administrative policy clearly stated: **Institutional data** are a **strategic asset of the University**. This data governance policy is based on the philosophy that **the value of data** as an institutional asset **increases with availability and appropriate use**. In order to support access and appropriate use, data must be managed using sound data governance principles.

Current Data Governance Landscape

Shortly after the data governance policy was released, a data stewardship model was rolled out to the identified stewards. The rollout was necessitated by a

specific enterprise data issue that needed to be communicated. The names of a few colleges were changing, having a ripple effect across several systems. The newly named stewards were excited to know that a stewardship organization had been created for facilitating enterprise-wide data conversations.

The data steward model is still being ironed out and expanded to include more subject matter. Initial efforts did little to distinguish master from transactional data domains and stewardship. Updates to the model will consider these concepts.

The EIM Council has already identified the next steps for the program to take. First, work has begun to build an information catalog so that data users across the university can find the data they need. The team anticipates that this catalog will have several components:

- A web catalog giving users the ability to browse and search;
- Metadata management infrastructure that will include a data dictionary showing the authoritative source for definitions and data lineage information to document where data are, where they came from, and where they go;
- A request and approval workflow process so data consumers can make requests for data that can then be approved by data stewards; and finally
- A data delivery hub to manage the data delivery process.

Other future work they have planned is to further build their data governance community and begin to work on master data management.

Lessons Learned

- Research helps, but there is no one right answer.
- Pain points are also opportunities.
- Leverage the strengths of your organization.
- Don't let technology drive the process but don't forget it can be helpful.
- Beware of boiling the ocean – it is a time and resource draining approach.

General Advice

- Emphasize that DG seeks to formalize and standardize tasks that many people are already doing anyway.
- Find the people in your organization that are trying to do things with data but can't. They'll see the value of EIM and DG and will help build support for your program.

¹ See <https://www.isixsigma.com/dictionary/pugh-matrix/> for more details.

Detailed Case Study

Introduction

This detailed use case is from one of the largest private, nonprofit integrated health systems (both provider and payer) in a major US city. The company started in the early 1900s as one hospital and clinic, expanding over time through mergers and acquisitions with other local hospitals, clinics/physician groups, ambulatory facilities (such as surgery centers), and postacute facilities (like skilled nursing facilities). The company operates or has built institutes for dedicated specialty care (including co-branding arrangements with other organizations to provide advanced care for specific specialties), a chain of urgent care and walk-in clinics, and various ancillaries including home health, laboratories, imaging centers, surgery centers, endoscopy centers, and pharmacies. Additionally, the company operates one-third of the city's emergency departments and trauma centers and often partners with state and federal government agencies on disaster preparedness (including the deployment of medical response teams for local, state, federal, and international events).

The company is fully licensed by the state to operate a variety of health plans including an employee health plan, fee-for-service plans, and co-branded plans with major carriers in the state-based marketplace and is now moving into the value-based care arena with accountable care organizations. This large variety of corporate functions, auxiliaries, subsidiaries, and co-managed services results in a very decentralized and highly fragmented enterprise which can be a difficult environment in which to introduce data governance (DG).

Organizational Background

The company is led by a cross-functional board of directors, an executive team that defines enterprise strategic goals and initiatives, and a corporate management team that oversees division and subdivision operations and objectives to support strategic goals. Few organizational charts exist, and they are hard to find and not always up to date. Because of the conglomerative-like structure of the organization and the highly regulated industry in which it operates, decision-making is slow and cautious, and several historically decentralized divisions/subdivisions operate with their own silos of people, processes, technology, and data. As is often the case with large, organically grown, and diverse organizations, territoriality, duplication of effort, and manual/semi-manual processes (often based on outdated legacy procedures) are rampant.

Hospitals function as discrete entities and leverage on-site administrators who interface with corporate administration. Physicians and surgeons are not directly employed by the company and managed through medical group models which also function as distinct entities with their own administrators (who also liaison with corporate administration). Co-branding arrangements are jointly managed. Fiscal-year planning sets formal enterprise goals that consistently revolve around revenue, cost, workforce engagement, patient satisfaction, and physician engagement; in more recent years, those goals have expanded to include health plan enrollment and population health management. Strategic plans are not always accessible or updated in a timely fashion, and leadership is exploring ways to better cascade strategically aligned goals and objectives across the enterprise.

Financial and budget management goals remain a top priority, often generating new profit improvement and Lean process reengineering objectives during each fiscal year. Limitations of aging legacy technologies coupled with assessments from various consulting firms specializing in health-care technology, process improvement, and strategy elicited a decision to modernize the company's electronic health record (EHR) platform for clinical care and revenue management, leveraging its implementation as a transformative catalyst to drive operational efficiency, eliminate waste and variation, and strengthen its trajectory toward a national health system.

Culture

The company is perceived as the city's leading health system with an esteemed

reputation that positions it as a premium brand. The company consistently wins various awards (Best Hospitals, Top Clinical Service Line, Best Company, Best Company for [Role/Attribute], Workplace of the Year, etc.) from Truven, US News, Gold Plus, and Advisory Board. Some physicians in specific clinical specialties are nationally recognized experts.

Leadership carefully protects its brand image, which is not always matched by internal operational efficiencies as the company does not yet function as a coordinated and integrated enterprise. Culture is heavily relationships-based, which yields slow and cautious decision-making complicated by decision revisitation as new stakeholders become aware of agreements through trusted personal communication channels that often supplement or replace formal communication channels.

Numerous councils, committees, work groups, and other team formats abound; charters are difficult to find; team purview is unclear or overlapping; operational decisions are poorly documented. Tenure is valued, with many employees on the higher end and many senior leaders in the latter stages of their careers. Organizational restructuring occasionally yields new coverage boundaries (including a few new CxO positions). Aging staff and skills are carefully being modernized through training, contract labor, and recruitment; attrition is leveraged for cost management to balance higher compensation levels needed for newer skill sets.

Finance areas have for years been the foundational driver for the company, though the transition to value-based care is increasingly emphasizing the need for clinical operations to enhance care delivery and patient outcomes across all care venues and locations through standardization and optimization. Information technology (IT) has historically been understaffed and inconsistently perceived as a critical central player; many decentralized divisions and subdivisions have embedded IT positions and functions into their business units, with varying levels of maturity and consistency of IT standards and processes.

Company headquarters span multiple geodistant locations, which complicate collaboration (although video conferencing and telecommuting are increasingly pervasive). Email is heavily used for communication and documentation. Content management standards, processes, and procedures are strong in clinical areas, improving in technology areas, and challenging in many business areas. Tribal knowledge is highly exploited and frequently lost with employee turnover.

Pay rates for IT staff are known to be lower than both other industries and local competitors, which contributes to turnover and creates ongoing challenges

with recruitment and retention. Staff bonus programs are based on financial and patient satisfaction goals, with corporate rates calculated as an average of all locations (thus rarely reaching maximum payout). Management bonus programs are based on attainment of financial and strategic goals and offer higher payout than staff bonuses. The company provides many nonfinancial benefits (e.g., alternate work schedules, remote work, discounted entertainment offers, and employee wellness program rewards that discount health insurance premiums). Paid time off seems less generous than other industries and some competitors.

Employees function with a “family” mentality, where tenured employees embrace brand loyalty and generally everyone is willing to pitch in and help each other with whatever is needed. That said, the company does have the requisite share of disgruntled personalities and those who engage minimum effort. Overall, most view the company as a favorable place to work where camaraderie and collaborative learning counter inequities.

Need for Data Governance Recognized

The first need for DG identified was the absence of a formal framework to address data issues with the existing enterprise data warehouse (EDW). The strategy during this time was to find an executive sponsor for the data warehouse to create accountability for data decisions and start managing the rampant shadow IT solutions and rogue technology tool purchases perpetuating that shadow IT.

Early on, the internal efforts to build a data warehouse focused on creating a central data platform of groomed data for analytics, but the lack of a strategic architecture plan or design yielded a collection of fragmented point solutions built by different contractors over time to target specific business problems. Technical teams worked in relative silos with some touchpoints for overall status/communication, but no effective coordination for architecture and development standards. Daisy-chained data processing jobs with immature error handling often caused unnecessary cascading failures and solution outages. Business processes around project management and IT development were immature and inconsistently followed. Stakeholders were often overwhelmed with detailed technology requirements, moving delivery dates, and constant scope change.

Data transformation standards and templates were being established via consensus through a “standards team” but some developers refused to follow

them, and a lack of architecture oversight and accountability for following standards tacitly condoned that approach. (For example, one developer coded everything in complex stored procedures that other developers could not support, and often spent weekends completely rewriting his solutions to overcome defects or add enhancements.) Developers were forced to support their own solutions after deployment; cross-training was difficult and in many cases non-existent; and there was no clear separation of duties between development and production support.

The vice president (VP) over the corporate project management office was the initial champion for the first iteration of the data warehouse (EDW 1.0), and the centralized analysts under her were the primary customers of this platform. They were often frustrated with and didn't trust the EDW platform due to myriad data quality problems (e.g., missing or incorrect data, no formal definitions, variation in metrics and dimensions), inconsistent solution availability (including data refresh rates), and lack of end user support documentation. No formal DG was in place at this time. Master data, metadata, and data quality processes and standards were missing. Analytics and reporting standards were inconsistent or not used. Analysts often extracted data directly from both source systems and the EDW to integrate and massage in their own data silos, expanding the use of and trust in their own solutions with technology readily available to them (primarily, Microsoft Excel and Access). Duplication, variation, and manual or semi-manual processing were common; analytics resources lacked intermediate and advanced data processing and automation skills; and their solutions were not easily supported, scaled, or enhanced.

A second indicator that DG was needed involved the rearchitecture of the data warehouse platform (EDW 2.0) and the search for an executive sponsor to champion the value of applying DG to that new platform. There was an understanding that DG needed to be a key component of managing data on the new EDW platform to avoid the data issues prevalent in the first generation of the warehouse. Initial efforts to find C-level sponsorship for DG in the IT or compliance areas yielded no results, so the strategy shifted toward evaluating other executive candidates and looking for success stories from other health-care companies. The initial sponsoring VP was facing increasing difficulty meeting leadership's analytics needs due to limited resources as staff struggled with the increasing demands of maintaining their shadow IT solutions and their inability to scale up and out. Learning engagements with leading health-care competitors with more data warehouse and analytics maturity reinforced the need to

rearchitect the current EDW and address core technological and data deficiencies using DG.

As part of this EDW rearchitecture, the company engaged two consultants to assist with building the second-generation warehouse and tweak the existing first-generation warehouse to enhance its usability as an interim solution. Several months into that engagement, the company recruited a DG specialist with EDW implementation experience to assist with project management, business analysis, development standards, and technical support of the new EDW platform. While not in an official DG role, the specialist immediately began to champion strategic data architecture, development standards, DG, and formal project management using agile (Scrum) with smaller milestones and incremental, iterative work sequences to deliver value, manage scope, and provide better project and timeline tracking.

For the next two years, the consultants expanded their development staff and devoted efforts to building a data warehouse following a very particular (highly normalized) architecture to integrate a half dozen key data sources. They also started building a real-time operational data store (ODS) to drive real-time data into operational workflows at the point of care, even though no business units, stakeholders, or customers had requested it. All this activity was managed by the principal consultant (architect) who verbally coordinated work across a few dozen developers as one giant project with little to no documented work breakdown structures, timelines, or milestones. Business partners and customers received no incremental solutions or value and were promised a full-blown, mature EDW platform at the end of a two-year build effort.

The approach these consultants engaged caused turnover within the project. Suggestions being proposed by the newly hired DG expert were met with resistance, and business owners were not being tapped for data-related decisions. Attempts to operationalize code deployments across development, test, and production environments often failed, and the consultant routinely blamed staff skill sets and a dysfunctional culture. Staff grew restless with the approach, and customers were increasingly frustrated by the lack of deliverables.

Meanwhile, the DG expert focused efforts on project management, trying to deconstruct the warehouse work to create a program with more manageable work streams and clearly defined scopes, using accepted project management approaches and agile software development principles and tools. He was given responsibility for overseeing project managers and business analysts, many of whom assumed those roles without formal training or experience. He coached

them on project management methodology and the core principles of DG and data management so they could factor them into their teams and projects.

In addition, the DG expert conducted research and attended a variety of data management and DG conferences where he learned about and then employed health-care analytics best practices to inform the data warehouse and business intelligence strategy and reinforce the need for DG. He also was able to contract with a health-care professional with decades of relevant experience who had launched a successful DG program in tandem with their EDW platform and had begun providing DG advisory services around her model. The company began working with her to educate leadership and the analytics community on the meaning of DG and the elements of a DG program they needed to implement. However, they still weren't "doing" DG and didn't yet have an executive sponsor for the program, nor anyone to help guide the data warehouse development and strategy despite many active attempts to find one. Through the collaboration and awareness gained at conferences, the DG specialist learned that health care was significantly behind other industries, and few organizations were maturing in the DG space.

Although no formal data roles or data decision boundaries were defined for the new EDW platform, they did map subject areas (roughly equivalent to data domains) and engaged a few business managers to help with data integration rules and data quality metrics within the new EDW operational processes. Still no real evolution of DG was occurring. Processes to capture metadata were scarce; data elements, metrics, and business glossary terms were not being defined, documented, or assigned to stewards; and formal taxonomies did not exist and weren't being defined. But the persistence was starting to pay off: even though formal DG was still neither sanctioned nor understood, they were consistently communicating about work being done, and the steering committee overseeing the data warehouse development was finally starting to understand some of the basic DG concepts and how they could be leveraged to solve the data challenges that were surfacing.

The third piece of evidence for the need for DG was the decision by senior leaders to upgrade to a modern EHR system. The DG specialist had already shifted his communication strategy to 30-second elevator pitches to convey key messages to leadership around changing internal data management processes. These sound bites included targeted points like *"Without data governance, we are swapping several dozen piles of garbage in our legacy applications for one big pile of garbage in the new EHR"* and *"Data governance will help us realize*

the value of our half a billion dollar investment". Messaging also included information to spotlight the siloed business areas where DG principles and processes were already being applied (especially to fulfill externally defined regulatory requirements like Finance and Accounting) and to position DG as the vehicle to formalize those behaviors into one integrated DG program across the enterprise. Health-care organizations were starting to appear at information management and DG conferences, and the DG specialist used lessons learned from those forums to reinforce internal DG messaging.

As part of the decision-making process for the new EHR implementation, leadership engaged professional consulting firms to evaluate their business model against industry norms and benchmark against competitors. In addition, a new physician leader contracted a separate evaluation with an independent professional. All those assessments supported the need for an integrated data warehouse architecture and called out gaps in master data management, metadata management, data quality, DG, formal project management, and production support. Leadership was especially attracted to the modern EHR platform's analytics platform which included both relational and dimensional data warehouses; several thousand pre-built reports, dashboards, and metrics; and dozens of self-service data marts. They saw this robust platform as a solution for their data warehouse problems, falsely believing the EHR analytics platform would solve all their data problems. The DG communication plan was quickly amended to add more elevator pitches around *"People and processes before tools!"* and *"Technology does not solve people and process problems."* to reinforce the need for DG with the new EHR.

Deployments with the company's rearchitected data warehouse (EDW 2.0) were still fragile. Internal technology architecture reviews and cost analyses indicated the best option going forward was the adoption of a data lake as an additional data storage architecture to complement the EHR's mature data warehousing/analytics platform and also provide data archiving services for data from legacy applications being replaced by the new EHR.

Around that same time, the existing Chief Information Officer (CIO) retired and the company brought in a new CIO experienced in that EHR to oversee the implementation. He supported the strategy of using the EHR's data warehousing platform with a data lake and demanded an inventory of all reporting and metrics with clearly assigned ownership and sign-off prior to go-live to avoid problems he faced with prior EHR implementations.

At this point, all efforts to initiate DG had reached a point of convergence.

The DG specialist began inventorying all metrics, reports/dashboards, and extracts for the new EHR and defining the data domains that would be used to assign ownership and stewardship to those assets. He also worked with internal privacy, compliance, and information security staff to standardize data roles, establish guidelines around data exporting, and address concerns about managing the extensive data movement with shadow IT solutions.

After the new EHR went live, various business units began clamoring to get broader access to its data so they could expand their own analytics solutions and incorporate some of the EHR platform's clinical and revenue cycle data. Additionally, some business units began hiring data scientists to support population health management initiatives and bring the company closer to the predictive modeling and advanced analytics needed to support personalized medicine. Internal conversations around an enterprise strategy for advanced analytics and artificial intelligence surfaced, and leadership resurrected prior conversations around how best to drive consistency of metrics and manage the hundreds of analytics resources sprinkled throughout the organization (all of whom were decentralized and embedded within business functions, except for the centralized IS reporting team supporting all the new EHR reporting and analytics).

Leadership also began talking about how to address variation in output from the various analytics teams, leading the CIO to launch a Lean initiative (sponsored by the CEO) to standardize analytics and reporting across the enterprise within the framework of the operational governance structure and DG program implemented with the new EHR. That effort initiated an enterprise-wide survey to identify analysts, measure the time they spent on various analysis and reporting activities, and inventory their data sources, skills, solutions, and customers. Insights from analyzing that data were used to inform an ideal (future) state of analytics and reporting and the interim steps to move toward that state (e.g., eliminate duplication/overlap and inconsistency, build an analytics community, and elevate analytics capabilities). The first concrete result of the survey analysis led to the creation of an Enterprise Analytics Steering Committee (see *Data Governance Organization – Timeline/2019: Current State [post EHR implementation]* for more detail) to collect and prioritize all analytics and reporting requests and manage the analysts fulfilling those requests.

As the use of the new EHR matured, the IT director of enterprise analytics and reporting set a goal to greatly mature their analytics capability by following the “HIMSS Analytics Adoption Model for Analytics Maturity (AMAM)” and

achieving the highest maturity level (Stage 7) within two years. ¹ Fortunately, the DG expert was already using a remarkably similar model (the “Health Catalyst Analytics Adoption Model”) to guide the evolution of their Data Governance Program; ² he is now mapping DG program deliverables to the HIMSS model (which carries nationwide prestige) to better prioritize DG program deliverables to support the goal of reaching HIMSS Stage 7 certification.

Data Governance Organization – Timeline

2012: First-generation Data Warehouse (EDW 1.0)

There really was no data governance office (DGO) or organization. The data warehouse steering committee meetings at the time were dysfunctional and ineffective; cultural challenges revolved around the politics and problems associated with the fragmented point solutions, and efforts to revamp committee membership, purview and focus consistently failed.

The executive sponsor for the first EDW implementation was the Chief Medical Officer; when he left the company, responsibility for the data warehouse work transitioned to the CIO because the data warehouse platform and point solutions were IT projects. The CIO insisted on finding a business (non-IT) executive to assume sponsorship and evangelize the need for a central platform of groomed data, and he targeted the Chief Compliance Officer as the best candidate with the strategy of addressing the shadow IT risks to compel the migration of those hidden information factories to the central platform as part of the next phase of warehouse development. This approach clearly reflected the same maturation process outlined in the Health Catalyst Healthcare Analytics Adoption Model which they were already following.

2013–2014: Second-generation Data Warehouse (EDW 2.0)

In 2013 and 2014, the newly rearchitected data warehouse was implemented, leveraging some business managers to act as informal data owners for specific subject areas. The DG expert allocated a third of his time toward educating

others about DG, in effect creating an informal DGO with one-third of a staff member (0.33 Full Time Equivalent, or FTE).

In addition, the EDW management team and the DG expert began meeting with the Chief Compliance Officer (a tenured executive with political capital) to enlist his support for DG, but the sponsorship question would not be so easily resolved. The Chief Compliance Officer didn't view the compliance department as being the right fit to drive changes in data management behaviors; he believed that the CIO should own DG along with the data warehouse. At that point, the DG expert created another 30-second sound bite (*The business owns the data.*) and insisted the business must drive DG with C-level support.

Fortunately, both executives were members of the newly revamped data warehouse steering committee which had been pared down from the days of EDW 1.0 to a small number C-suite and VP-level members. That team was named the collective sponsor of efforts to rearchitect the EDW and launch DG, and members were refocused on tracking the status of the data warehouse rearchitecture and managing incoming requests for analytics, approving and prioritizing only those requests that were truly strategically aligned and necessary. Eventually, the committee went on hiatus as the work focused solely on the data warehouse development work and no new analytics requests were entertained. In lieu of the committee, the DG expert met on an ad hoc basis with the new Chief Medical Officer and the Senior VP for Strategy (who had agreed to be chair of the steering committee) to address approval and prioritization needs that still trickled in, until a formal blackout for new analytics requests was implemented.

2015–2016: EHR Preparation

In 2015, the new VP for IT hired to assist with the EHR implementation that initiated an organizational restructure that yielded a set of new “team lead” positions for specific departments. He did not fully embrace the role of DG, and no formal DG role was created; the DG specialist was reassigned to the team lead role for the department implementing the new EHR's data warehouse and analytics platform.

Later in the reorganization process, an employee with analytics and reporting experience proposed, and was given, a new “Data Governance Lead” role to drive DG on the new EHR platform, yielding an informal DGO with 1.0 FTE staff. This employee spent time working for the next 18 months with business

and IT staff on the EHR preparation phase: collecting requirements, identifying owners for the EHR reports, and working with the DG advisor on a plan to launch a formal Data Governance Council (as successor to the former data warehouse steering committee) and Data Governance Program to govern the EHR reporting and analytics content.

2017–2018: EHR Implementation (EDW 3.0)

About nine months before the new EHR was due to launch, the official Data Governance Lead left the company. DG duties were reassigned to the data warehouse manager and the previous DG expert. It quickly became apparent that the EHR implementation needs did not yield enough capacity for the data warehouse manager to assist with DG, so the DG expert informally assumed the Data Governance Lead duties using about 70% of his time (0.7 FTE capacity) with a half-time (0.5 FTE capacity) physician co-manager to assist with DG tasks and ongoing education with leadership and management.

They were still in need of an executive sponsor and solicited the support of the Senior VP for Strategy because of her background as an analyst and past assistance with addressing data warehouse priorities. She was also supportive of DG efforts and accepted the request to become the business chair of the new Data Governance Council. The VP for IT over the EHR implementation was named co-chair to ensure a strong IT–BIZ partnership.

They reformulated the Data Governance Council into a team of nine leaders with purview over both the EHR data warehouse and analytics platform and the Data Governance Program, and launched two cross-functional DG work groups: one to standardize reporting security roles and establish data policies and standards for the new EHR platform, and one to address requests for new (or changes to existing) EHR reports, metrics, and extracts. This structure is shown in [Fig. 12](#).

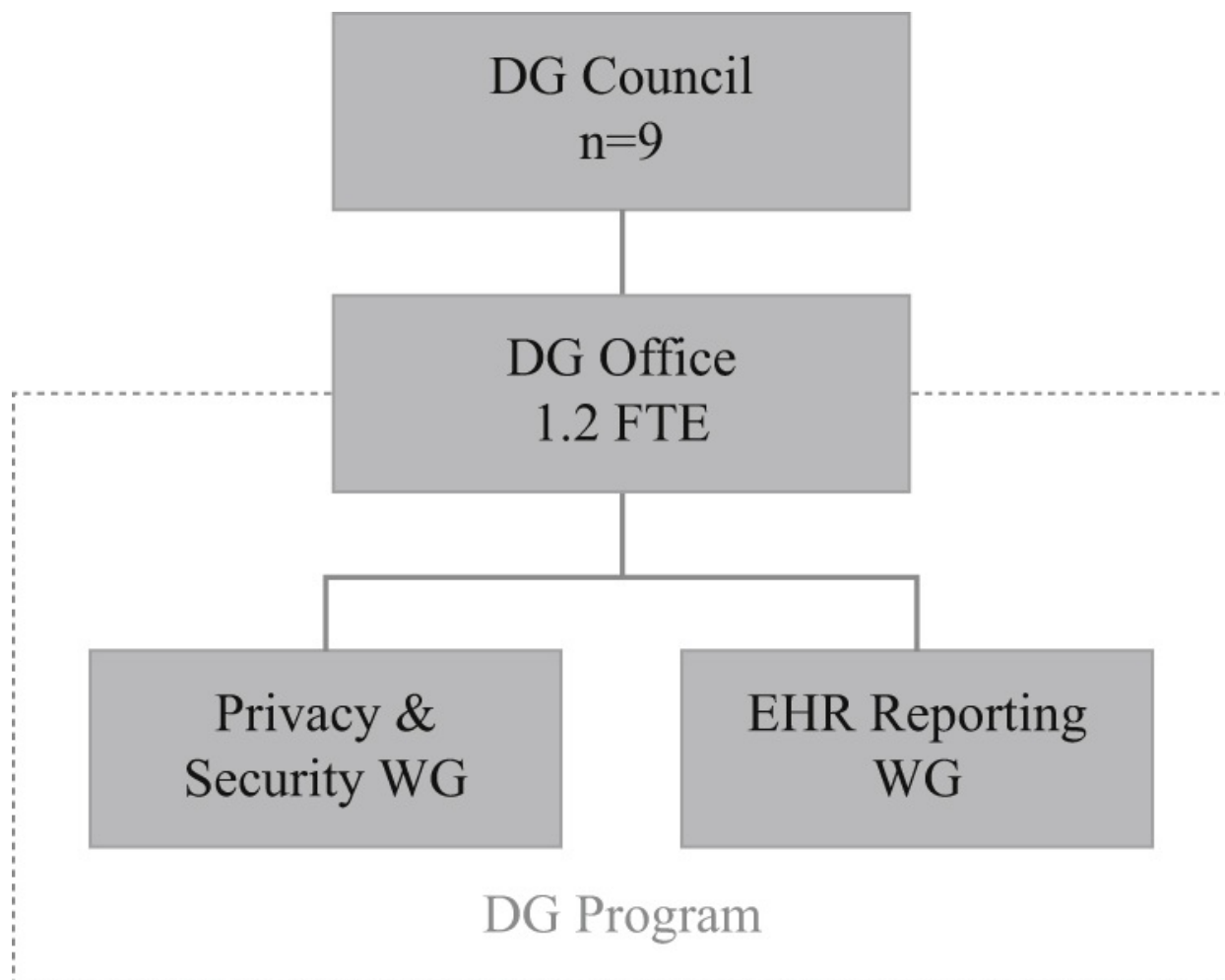


Fig. 12. First Data Governance Structure.

At this point, the existing CIO was nearing retirement and the new CIO for the EHR implementation had already started onboarding. With his mandate to inventory reports and metrics, assign owners, and standardize security roles for the new EHR, he was quickly recruited to become the IT co-chair of the DG Council to support the DG program to meet those goals.

Working with the DG advisor, the DG expert proposed 83 data domains each with a dedicated Information Owner responsible for approving data-related requests (mostly for new reports/extracts/metrics or enhancements to existing reports/extracts/metrics) and solicited DG Council endorsement for those roles. Reports and extracts were then mapped to a primary data domain based on either (1) which data domain best aligned to the majority of data on the report or (2) which business unit was the primary user of that data for daily operations.

Metrics were mapped to one data domain.

2019: Current State (Post EHR Implementation)

By the end of the two-year-long EHR implementation, the DG expert became the full-time Data Governance Program Manager (1.0 FTE capacity) with an official DGO and Data Governance Program. An additional internal resource was assigned to the DG Office to assist part time (0.2 FTE capacity), yielding an official DGO with 1.2 FTE.

As part of the Lean initiative to drive analytics consistency, the EHR Reporting work group was reformulated as an Enterprise Analytics Steering Committee with purview over demand management (analytics requests), supply management (analyst resources), and prioritization and execution of analytics initiatives enterprise-wide. The DG Office holds purview over the Data Governance Program, and they are in the process of launching new workgroups to support each of the DG workstreams as they are activated. The Data Governance Council oversees both the steering committee and the DG Office.

The DG Office is internally recruiting existing BIZ and IT staff to spend part of their time helping manage DG workstreams with the aim of growing into full-time Data Governance Practice Leads if they are able to get approval for additional headcount (obtaining new positions is a cultural challenge). This proposed structure is shown in [Fig. 13](#).

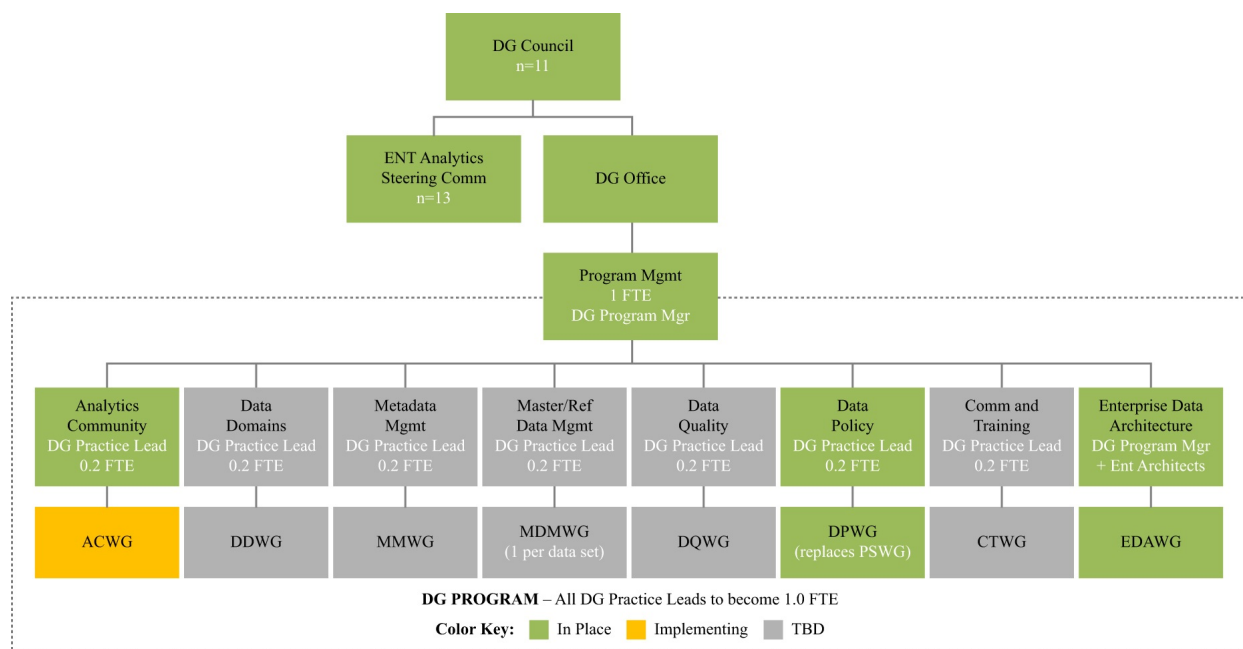


Fig. 13. New Data Governance Structure.

The current Data Governance Program is organized into nine workstreams:

1. **Analytics Community:** Establishing standards for analytics development; providing skills and training for analysts; elevating analytics competencies.
2. **Data Domains:** Defining official data decision-making boundaries; assigning owners and stewards; mapping data domains to strategic goals.
3. **Data Quality:** Establishing data quality standards and processes; identifying and remediating data defects; leveraging data quality reports and dashboards to build data trust.
4. **Metadata Management:** Cataloging data assets; mapping technical metadata and lineage; standardizing taxonomies to tag assets for searchability; and enriching the catalog with business metadata.
5. **Master Data Management:** Implementing master data management for business-critical shared data sets (11 master data domains); managing reference data; driving consistent use of these data sets across all analytics and reporting.
6. **Data Policy:** Establishing data-related policies for different stages of the data life cycle; defining data privacy, security, and auditing controls; standardizing data roles and permissions.
7. **Communication and Training:** Marketing and public relations about the DG program; formal training for the DG program, roles, data sources, and analytics tools.
8. **Data Architecture:** Standards and tools for centralized analytics platform; connecting data architecture strategy to infrastructure teams and overall digital strategy.
9. **Program Management:** Managing DGO operations; coordinating DG work group meetings; facilitating intake and triage for DG and data

challenges; providing data management consulting; overseeing the DG program.

The Senior VP for Strategy and the CIO remained the executive sponsors through the EHR implementation. However, many mature health-care organizations leverage the role of Chief Medical Information Officer (CMIO) to drive optimization efforts around analytics, reporting, EHR workflows, and DG to improve clinical operations, efficiencies, and outcomes. During the ramp up for the new EHR and the expansion into value-based care, company leadership initiated recruiting efforts for a CMIO but did not succeed in finding the ideal candidate. Recently, after the initial EHR implementation concluded, two internal physicians who worked closely with IT on the EHR implementation were named CMIOs with one dedicated to ambulatory care and the other focusing on acute care. The CMIO roles may eventually become candidates for joining the Data Governance Council to sponsor the DG program and drive efforts toward DG goals.

Getting Started

In the early days, the primary task at hand for the first-generation data warehouse was educating everyone (especially leadership, management, and analysts) on the purpose, value, and basic tenets of “good data” as an organizational strategy. Messaging focused on the following points:

- Manage data as an asset, with official owners and stewards.
- Consolidate analytics and eliminate multiple warehouses and shadow IT solutions.
- Reduce the number of business intelligence tools and corral rogue IT purchases.
- Provide consistent messaging through 30-second elevator pitches.
- Learn from health-care companies with more mature data warehousing and analytics capabilities.

Moving to the second-generation data warehouse focused on messaging

around continued education and introducing the idea of subject areas for identifying owners. Key principles to expand understanding of DG included:

- Position “groomed data” as the path to better analytics.
- Drive toward common definitions (metrics, dimensions, and business terms) with clear ownership.
- Move key analytics into the data warehouse so everyone benefits.
- Measure our analytics footprint (analysts, data sources, solutions, tools, skills) and look for opportunities to eliminate redundancy, overlap, and variation.
- Modernize our data architecture especially around master and reference data, metadata management, and data quality (automating as much data processing as possible).
- Formulate and evangelize an enterprise strategy for reporting and analytics.
- Enlist the expertise and assistance of Internal Audit and Compliance to address data privacy and security risks with legacy behaviors bleeding into the new EHR platform.
- Leverage research, conference insights, and our DG advisor to evangelize the role of DG as a core business function.

The next wave of change brought about by the new EHR implementation required yet another set of steps to evolve everyone's understanding of DG:

- Map the impact of health-care industry reforms to the need for better data.
- Link analytics needs and roadmaps to our strategic plan and major initiatives.
- Leverage external consulting assessments not only to inform the EHR implementation but also to position DG as an enterprise strategy.
- Formulate a master data architecture strategy for the new EHR with consideration for related “big data” issues.

- Collaborate with our DG advisor to build out our DG strategy and plan.
- Establish a DG program to coordinate these activities and a Data Governance Council for oversight and decision-making to fulfill the new CIO's goals for the EHR implementation.
- Inventory EHR reports and metrics with clear definitions and assign them to data domains and owners.
- Standardize reporting security roles and associated functionality privileges for the new EHR and map all users to a role.
- Implement a centralized report and metric request process and a governance team to vet, approve, and prioritize those requests.

And in the final, post-EHR phase, there was still more work to be done:

- Create awareness of the DG program and expand understanding of DG.
- Provide assistance and consulting support to users experiencing data quality challenges or analytics and reporting issues.
- Develop a process to triage data challenges and questions to the right DG program workstream for resolution (or incorporation into the DG program roadmap), reroute to the appropriate business or IT processes, or escalate to the Data Governance Council for remediation.
- Inventory master data sets and needs.
- Engage analysts to evaluate and select an enterprise data visualization tool as a standard.
- Compare and contrast analytics team organizational structures (centralized, decentralized, hybrid) to evaluate pros/cons and cultural drivers and challenges.
- Reformulate the EHR Reporting work group to an enterprise analytics steering committee with purview over all enterprise analytics and reporting.

Moving on from all the technology-focused work to operationalize DG, the

organization still has the following next steps they plan to address:

- Build a DG Office staffing roadmap with clearly defined roles and responsibilities for DG Practice Leads to manage DG workstreams and workgroups, set key milestones for workgroup deliverables, and facilitate stewardship communities.
- Engage DG Council members to launch workstreams assigned to them as sponsors.
- Recruit staff for DG Practice Lead roles.
- Build master data business cases around provider, location, and identity data to support a capital funding request for development resources and tools.
- Evaluate and select a metadata/catalog tool and initiate cataloging.
- Designate Analytics Stewards and launch the Analytics Stewardship Community to drive analytics development standards and processes.
- Stabilize the Enterprise Analytics Steering Committee and implement an intake process and prioritization algorithm for all analytics requests.
- Engage analysts to continue inventorying current analytics solutions.
- Build an intake mechanism to collect and triage data-related challenges to the appropriate DG workstream.
- Add several hundred EHR and corporate business terms to the business glossary.

Results

- Formal inventories of data assets with metatags and owners:

Inventory		Contents	# Records
Report Tracker		EHR reports, dashboards, scorecards	~3,100
Metric Tracker		EHR metrics and KPIs	~2,600

Extract Tracker	Data extracts from EHR data warehouse	~350
Data Mart Tracker	Cubes, data models and universes for self-service analytics	~100
Database Object Tracker	Database-level variables used across the EHR platform	~500
Databases	SQL instances across the enterprise	~20,000
Glossary	Data governance terms and acronyms	30
Application Tracker	Applications and software used across the enterprise	~1,200
Analytics Inventory	Self-service analytics built by BIZ analytics teams	~700
Analytics Teams	Self-service analytics team names, members, manager, and customers	23
Data Domains	Official data-related decision boundaries and assigned Information Owners for both the EHR and ServiceNow implementations	131

- Eight official DG program roles defined and implemented: Information Owner, Information Steward, Report Owner, Analytics Steward, Application Owner, Application Steward, DG Program Manager, DG Practice Lead.
- Five standardized reporting security roles for EHR access/privileges for all employees (ranging from read-only data consumer to full-blown data developer).
- Nine EHR data-related access and usage standards (which are not official policies yet).
- Approximately 400 EHR requests for reports, metrics, extracts.
- Provider MDM pilot integrating credentialing, health plan, and HR data yielding golden record set of 1,415 provider records of 10 data fields with 99% automated match rate using five rules.
- Data profiling results on 24 fields in two source systems which showed true error rates (missing or inaccurate data), the need for data standards, and the risks of tribal knowledge.
- Increased communication and collaboration across business units to understand data collection and validation processes, data needs, and the

impact of their data decisions on others.

- Collaboration with revenue cycle staff to validate ~30 EHR metrics across 40 reports.
- Collaboration with hospital operations, finance and quality reporting staff to validate ~40 census metrics and definitions and clarify usage.
- Deconstructed eight data domains into multiple subdomains including the assignment of owners/stewards and mapping to EHR operational governance teams.
- Four conference presentations and one webinar about our DG Program structure.
- Thirty-five DG training classes for Information Owners and DG Council members.
- Established the DG Office with one DG Program Lead and a part-time DG Practice Lead.

Lessons Learned: What Helped with Success

- Engaging the right executives to sponsor and evangelize DG.
- Connecting executives to DG work to clarify priorities and amount of effort.
- Soliciting leadership support by presenting DG to senior executives with CEO backing.
- Embracing change as uncomfortable and being willing to be patient and help folks through it.
- Promoting transparency as DG's best friend.
- Allocating dedicated DG Office resources and protecting them from competing cultural challenges.
- Being persistent and knowing that one can NEVER GIVE UP because the only way to fail is QUIT.

Lessons Learned: What Got in the Way

- Setting strategic goals set without considering impact to (or the role of) DG.
- Lack of consistent enterprise-level prioritization and strategic alignment of work.
- Struggling with intake and prioritization of projects/enhancements.
- Having limited DG resources and capacity management skills/tools.
- Lack of funding for DG Office staffing and tools.

Lessons Learned: What to Do Differently

- Blend EHR vendor's suggested structure with our organizational structures to foster better adoption and accountability of operational governance team structures.
- Coach DG roles on data-related responsibilities and hold them accountable for doing their jobs.
- Avoid the tendency for DG Office to do the work for owners/stewards and DG Council.
- Plan DG goals and deliverables for each program phase up front and use status clear reporting and escalation to DG Council to keep everyone on track.
- Implement and manage all data asset trackers in the same tool and drive consistency of data fields and values.
- Recognize and support the need for enterprise-level master data management as a key data architecture strategy.
- Implement formal tracking and accountability for sign-off on metrics, reports, and extracts by information owners.
- Establish a change management process for data domains and information

owners with assigned accountability for maintenance up front (well before go-live).

- Provide DG training up front (prior to go-live) and certify owner/steward roles to make them more official.

Ongoing Challenges

- Cultivate understanding that DG succeeds best when the business pulls DG while the executives push it.
- Balance DG as a strategic function while navigating the organization's crisis-based firefighting mentality.
- Explain how to leverage DG as a vehicle to drive cultural transformation and eliminate shadow IT, data hoarding, and data territoriality.
- Devise formal DG Program metrics to communicate progress and success.
- Build trust in central data architecture and wean people off shadow IT.
- Help owners/stewards live up to their roles and responsibilities to overcome the cultural challenge of accountability.
- Compile a solid business case for metadata management and master data management so senior leaders understand them and are willing to fund the technology and processes to do them correctly.
- Elevate analytics capabilities and standards so analysts can excel with advanced analytics, machine learning, and artificial intelligence.
- Redesign the DG documentation and communication platform.

1 <https://www.himssanalytics.org/amam>.

2 <https://www.healthcatalyst.com/healthcare-analytics-adoption-model/>.

Execution Roadmap

It is evident from the chapters and case studies that data governance is messy business. It is not a standard formula or a neat and well-bounded algorithm that is guaranteed to produce results. Our goal in this book was to make this point emphatically so that business leaders don't expect immediate results or permanent successes. Data governance is not for the faint-hearted. It is for those who are patient and persistent in their attempts to pursue a somewhat slippery goal. But the results of data governance can be far-reaching for an organization and hence, it is well worth pursuing this goal. As the previous chapters and case studies have shown, extraordinary results and lasting benefits can be achieved by following the well-researched advice of experts and practitioners. This book covered several critical success factors that data governance advocates, activists, and implementers must take into account before starting or furthering a data governance program. We presented key ideas on important considerations and design issues that will make a data governance program successful in the long run. This chapter serves as a guiding document, a roadmap, and a critical checklist on how to accomplish those goals. The ideas presented here are universal, in the sense that almost all organizations and data governance advocates are likely to use these guidelines at some stage or the other in their data governance journey.

Key Points

1. **Define Data Governance:** Take the time to define what data governance means to the organization. There is no universal definition of data governance because a good definition should take into account the culture of the organization, its current state of affairs, its goals and

objectives, the nature and scope of its data needs, and its regulatory and compliance requirements, among other things. The more customized the definition of data governance is, the more it will resonate with the audience.

2. **An Ongoing Project:** Emphasize to all stakeholders that data governance is an ongoing journey and that patience and perseverance are the hallmarks of successful data governance initiatives. Data governance is not easy to execute because it demands short-term action and long-term focus. Outline the challenges inherent in this process and prepare stakeholders to meet these challenges.
3. **Not an IT Issue:** Reiterate that data governance is not an IT responsibility but is everyone's responsibility. Business and IT must work together to succeed. IT holds the technical end of the tight rope while business holds the competitive issues that will shape the organization. IT and business are the two wheels of the bicycle and both must move in a synchronous fashion if data governance is to succeed.
4. **The Three Cs:** Build the data governance strategy and implementation tactics around three pillars: change management, culture, and communications. If any one of these pillars is weak or neglected, the success of data governance will be compromised. Data governance leaders must keep their eye on all three pillars.
5. **Culture Is Key:** Understand the culture of the organization. This will yield immense rewards. Align data governance definitions, goals, and priorities to match the culture. Pay attention to the formal and informal culture that is embedded in the organization. Pay close attention to subcultures and the role of influencers in the organization. Select the initiatives carefully and customize it to align with unspoken values and priorities. We presented different cultural models. Review them and tweak them to fit the needs. An organization does not have to fit perfectly within any of the models we presented. The models are presented just to help find an amalgamation of different attributes from different models. But it is important to draw a framework that identifies the critical cultural attributes of the specific organization.
6. **Change Management:** Invest heavily in change management initiatives

in order to truly build a data-driven culture. Monitor the success of each initiative closely and identify the critical elements of change management. Tweak the model by noting lessons learned. Articulate, over and over again, to the stakeholders the benefits of being data-driven.

Lessons Learned from the Case Studies

How did our case study organizations proceed to move things forward? Given all the previously listed considerations that explain how differently things can be done in different organizations, how can we translate that to what other organizations can or should do? Here are some concrete steps for using the key takeaways and lessons learned that we have outlined.

- **Know the Dream:** First understand the expectations and desires that the organization has for starting a data governance initiative. If data governance is not perceived as important or valuable, then focus on the expectations and desires that the organization wants to achieve in the business. This is the tactical equivalent to tying the data strategy to the business strategy. Even in organizations that do not have a clearly articulated strategy, leadership will have made known what some of the priorities are and data governance projects can be tied to those.

As we saw in the case studies, this first step takes different forms. In several cases, the desire to have data available from across the organization in a single repository was a key driver for moving forward with data governance. For another, the spark came from the desire to have one business intelligence platform from which to do reporting. These are tactical goals to achieve a strategic end but identifying them and understanding how data governance supports or advances the achievement of that goal will help get initiatives underway. In other words, what is the secret sauce?

- **Acknowledge the Politics:** When we understand the culture of the organization and its many subtle nuances, we can identify how to work within the culture and politics of the organization. Because there is no one way or “correct” method to start data governance, there is no concrete checklist of a list of individuals who must be involved or the capacity in

which they should be involved. As we mentioned previously, communication is key and politics and culture will help dictate to whom one needs to communicate, how, what, and when. It will also inform the governance framework and the roles and responsibilities that the data governance structure needs. There is no “right” way to do it; but there is a way to make it work for a particular organization, although this may evolve by trial and error.

The case studies highlight how these issues play out. In one case there was an internal practice such as Six Sigma that was able to get things started and before including IT and doing an organization-wide “road show” to bring people on board. The culture and politics can lead to work being done on multiple levels with communication directed both from the top and from the bottom. Political wrangling can't always be avoided but it can be helpful to identify when it derails important work. In another case, the politics led to turnover and the “outside expert” bias wasn't recognized early which may have helped curb some issues. And in the case where the culture and politics contributes to viewing everything through a technology lens, it led to several restarts of data governance efforts.

- **Understand What's Broken:** Knowing where the organization wants to go isn't enough to get there. It's important to know what might get in the way of that journey. Roadmaps and GPS systems are more useful to drivers when they highlight accidents and construction that may delay or divert a traveler from getting to their destination. Data governance planning isn't very different. Discover what are the reasons that the dream has not been attained. It could be politics, broken processes, resource constraints, or a host of other issues. Remember, change is hard and regardless of the change management strategy or approach, getting to a goal requires something to change or the organization would have reached that goal already.

In each of the case studies, the organizations looked to identify what was getting in the way of achieving their goal. In some cases, it was issues with data sharing. In others, it was improving critical understanding of the issues that needed to be addressed. As we have said before, data governance is hard and not everyone in an organization who needs to understand that will understand that.

- **Identify What Can Be Changed:** After recognizing what is broken, it is important to identify what can be fixed. The navigation system in a car might identify the accident that is blocking the road, but any individual traveler may not be in a position to move the accident blocking the road out of the way. Some travelers may be able to choose a new route because that is something over which they have control. Similarly, some broken parts of a data infrastructure are not readily changeable so it is important to understand what can be fixed. If a key challenge is data literacy in the workforce, it's not reasonable to consider replacing the entire staff. It is reasonable to consider data literacy programs or training to address that gap.

In each of the case studies, those leading the data governance charge identified the avenue that would readily yield results. It may not have been optimal, but it was achievable. Some business processes can be adapted, some technologies can be obtained, and some education and introspection can be done to inform leadership. None of these activities are earth shattering, but each was attainable. And making some progress leads the way to more progress.

- **Draft a Plan to Change What Is Broken:** As was mentioned in one of the change management models, even the clearest vision with the best support cannot be achieved without a plan. Knowing where to go is only half the battle; we need a map to get there. Not all plans are created equal either. Some goals can be achieved with high-level objectives and a fungible timeline. Others need detailed project management with milestones and deliverables. Be as detailed as is needed, consulting the key people that need to be involved. In many cases, it will be more efficient to rely on a small core team to create a draft plan rather than trying to include every possible stakeholder from the beginning. As we have discussed at length, communication is the key and when to communicate with whom is going to depend on the culture and expectations of a particular organization. That said, a large group of people trying to create concrete items from abstract or amorphous concepts is a breeding ground for analysis paralysis: when overthinking and trying to plan for every possible contingency keeps a project from moving forward. A smaller team can create the example for a larger group to react which can be far more effective in creating a workable action

plan.

In each of the case studies, the initial work was done by a small team before collecting feedback from a broader stakeholder group. Whether using an existing business function, like Six Sigma, a cross-functional workgroup, or a team of consultants, each organization found a way to have a plan drafted for others to react to. Sometimes, the initial plan included upfront and explicit education before delving into details. If it is necessary to ensure a base level of understanding of data governance before the details of a plan can be understood and appreciated, make sure that information and education is built in. Otherwise, it will be hard for stakeholders to grasp both what the initiative is planning to achieve as well as how it is planning to achieve it.

- **Let Everyone Explain Why It's Wrong:** Generally known in business speak as “socializing” something, the garnering of feedback for any idea is both important and painful. As we mentioned before, change is hard and people try to avoid change that affects them personally but they will generally support change for someone else. Therefore, when meeting with stakeholders and interested parties about the plan for change, be prepared for a great deal of negativity. If it doesn't surface, that is great! Take the plan and run with it. However, a more likely scenario is a chorus of progress killing phrases such as “it's complicated,” “that won't work for us (we're special),” and the ubiquitous “we've always done it this way.”

The key at this stage is to make sure that everyone feels heard and acknowledged. Understand *what* they feel they are losing in the process of implementing data governance and see if there is a way to assuage that fear. Maybe additional education or training or reassurance is necessary or there is an erroneous assumption or conclusion. Maybe there is a real shortcoming and the plan needs to be altered. We suggest working with the clearly stated principle “everybody gets a say but not everybody gets their way”: the larger the stakeholder group, the more likely there is to be diametrically opposed opinions which cannot be appeased by a specific plan or action. This means someone will be disappointed. Acknowledging that up front and explaining why some suggestions are taken while others are not will help to smooth the path but it won't remove all the obstacles. We are also fans of the “80/20 rule”: if a plan works for 80% of the constituents, adopt it with the caveat that there needs to be additional work done to address the 20% for whom it doesn't work. In our

experience, the 20% often turns out to be closer to 5%: once people understand that they really aren't so special, they are often more able and willing to come on board. That said, sometimes it depends who is in the 5%. And some of those cases, politics will just get in the way but hopefully that has been taken into consideration (see point above).

In each of our case studies, there were many conversations during the planning process and not everything went smoothly. Or some went smoothly but not all issues were surfaced and so a restart was necessary. Or the plans needed to be adapted over time as work progressed and new issues arose. In each case, the recognition that change is uncomfortable was important to moving forward. Communicating with all interested parties and stakeholders at all stages, and in some instances including a formal communication plan, helped to inform the work at all three organizations even where there were concerns.

- **Do It Anyway:** We don't mean to sound flippant and we are not encouraging subordination. What we are encouraging is persistence. In many instances, the naysayers and complainers who object to the plan, either surreptitiously or vocally, will expect their objections to make the plan go away. And depending on the politics, that may be the most prudent course of action. However, if the plan really outlines how to remove the roadblocks to achieve an acknowledged beneficial goal, then it should be worth pursuing. There may be the need for many rounds of stakeholder discussions or education before a specific plan can be championed by a large group of constituents. That's okay. As we have reiterated repeatedly, data governance is hard. Maybe the way forward is to do a pilot project or “proof of concept” to help the skeptics clearly see the value in what is being proposed. People are often more willing to support change if they feel they can revert to the status quo after proving their objections were sound. Dangling the prospect of an “I told you so” moment can often sway many objectors and naysayers.

As the case studies show, and their recommendations clearly state, there will be missteps and restarts but perseverance is necessary. The “marathon, not a sprint” analogy is often heard and successful data governance practitioners regularly point out aphorisms about the journey of a thousand miles beginning with a single step. Their key is to find that small step to start the journey.

Conclusion

To do one's best against all odds and to be patient and to persevere is the mantra of the best data governance experts. They do this difficult work because they understand that in the long run, the organization and its stakeholders will benefit immensely. Data governance leaders are continuous learners – passionately seeking new knowledge while chopping off the branches of old and outdated information. They understand that data governance is not about technology or jargon, but it is about working with people. It is about collaboration and compromise. It is about building small bridges that then lead to bigger bridges that then lead to a better destination. We hope the journey is successful. We would love to hear about lessons learned so that we can collectively continue to advance the field of data governance. Good luck!

Notes

- i Cannon and Gupta (2020) explore this issue in greater detail.
- ii <https://hbr.org/2017/05/whats-your-data-strategy>.
- iii <https://www.prosci.com/resources/articles/what-is-change-management>.
- iv http://www.valuebasedmanagement.net/methods_quinn_competing_values_framework.html.
- v <https://www.ocai-online.com/about-the-Organizational-Culture-Assessment-Instrument-OCAI/Competing-Values-Framework>.
- vi <https://implementtoimpact.blog/2017/11/13/managing-complex-change/>.
- vii Oxford English dictionary.
- viii The Merriam Webster dictionary.
- ix <https://www.thrillist.com/eat/nation/names-for-sub-sandwiches-around-america>.
- x Point of clarification: For this discussion, a business glossary captures the definition of the concept being defined. This is not to be confused with a data dictionary which we will define as the technical definition of a particular instance of concept. The glossary describes the concept; the dictionary describes the actual data element that captures that concept. See <http://tdan.com/data-dictionary-vs-business-glossary/24652>.
- xi In research, the ability to appropriately describe and even cite a compendium of data is very important. Organizations with a research function would be well served to ensure that any definition of collective nouns for data elements works for both administrative and research functions. See https://www.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/240_Final_Submission.pdf.
- xii <https://www.thefreelibrary.com/The+Road+to+the+Data+Strategy+for+Army+Intelligence.-a0549160099>.
- xiii <https://www.fasttrac.org/blog/how-to-define-your-business-strategy/>.
- xiv Dallemule, L., & Davenport, T. H. (2017). What's your data strategy? *Harvard Business Review*, May–June.
- xv <https://www.cio.com/article/3234884/what-is-a-chief-data-officer.html>.
- xvi National Science Foundation recommendations for meeting data management plan requirement for funding. https://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#IIC2j.
- xvii Accuracy is just one dimension of data quality. See Chapter 6 for further discussion.
- xviii Valerie Logan. (2018, September 21). Information as a Second Language: Enabling Data Literacy for Digital Society. ID: G00365697.
- xix <https://www.dataversity.net/what-is-data-quality/>.
- xx See Sebastian-Coleman (2012), Appendix B for a review.
- xxi https://cdn.ymaws.com/edmcouncil.org/resource/resmgr/featured_documents/BP_DQ_Dimensions_Oct17.pdf. Accessed July 29, 2019.
- xxii Inspired by <https://www.infoworld.com/article/3259744/data-literacy-what-it-means-and-why-its-essential-for-your-business.html>.
- xxiii For example, see Babbie (2004).
- xxiv See <https://www.cdc.gov/tuskegee/timeline.htm>.
- xxv <https://www.sociologylens.net/topics/crime-and-deviance/laud-humphreys-tearoom-trade-the-best-and-worst-of-sociology/11418>.
- xxvi <http://eugdpr.org/the-regulation/gdpr-faqs/>.
- xxvii <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1027&context=ecis2011>.
- xxviii https://www.ibm.com/developerworks/community/blogs/bigdataanalytics/entry/big_data_analytics_maturity_model?lang=en

Resources and References

Suggested Resources

These works provide additional or alternative information to the content of this book.

Al-Ruithe, M. , Benkhelifa, E. , & Hameed, K. (2016). A conceptual framework for designing data governance for cloud computing. *Procedia Computer Science*, Elsevier B.V., 94(C), 160–167. doi:10.1016/j.procs.2016.08.025

Alhassan, I. , Sammon, D. , & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, 31(2), 300–316. doi:10.1108/JEIM-01-2017-0007

Bonvillain, N. (2019). *Language, culture, and communication: The meaning of messages*. Lanham, MD: Rowman & Littlefield.

Case Study: Crampton, J. (2015). Collect it all: National security, Big Data and governance. *GeoJournal*, 80(4), 519–531. Retrieved from <http://www.jstor.org/stable/44076314>

Dai, W. , Wardlaw, I. , Cui, Y. , Mehdi, K. , Li, Y. , & Long, J. (2016). Data profiling technology of data governance regarding Big Data: Review and rethinking. In S. Latifi (Ed.), *Information technology: New generations*. Advances in intelligent systems and computing (Vol. 448). Cham: Springer.

Delacroix, S. , & Lawrence, N. (2019, November). Bottom-up data trusts: Disturbing the ‘one size fits all’ approach to data governance. *International Data Privacy Law*, 9(4), 236–252. doi:10.1093/idpl/ipz014

Kamioka, T. , Luo, X. , & Tapanainen, T. (2016). An empirical investigation of data governance: The role of accountabilities. In *PACIS 2016 proceedings*. 29. Retrieved from <http://aisel.aisnet.org/pacis2016/2>

Khatri, V. , & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, 42(4), 303–312. doi:10.1177/0340035216672238

Moens, A. , & Roberts, N. (2017, March). Not just an IT issue: Why governance of data should be on the agenda of every board director [online]. *Governance Directions*, 69(2), 104–106.

Riggins, F. J. , & Klamm, B. K. (2017). Data governance case at KrauseMcMahon LLP in an era of self-service BI and Big Data. *Journal of Accounting Education*, Elsevier, 38(C), 23–36.

Seiner, R. S. (2014). *Non-invasive data governance: The path of least resistance and greatest success*. Basking Ridge, NJ: Technics Publications.

Srivastava, S. B. , & Goldberg, A. (2017). Language as a window into culture. *California Management Review*, 60(1), 56–69. doi:10.1177/0008125617731781

References

Ambrose, D. (1987). *Managing complex change*. Pittsburgh, PA: Enterprise Group.

Babbie, E. (2004). Laud Humphreys and research ethics. *International Journal of Sociology & Social Policy*, 24(3–5), 12–19. doi:10.1108/01443330410790849

Becker, T. (2019, October). *Discussing data: Evaluating data quality* . Technical Briefing no. 19-03. Federal Reserve Bank of Kansas City.

Cameron, K. S. , & Quinn, R. E. (2011). *Diagnosing and changing organizational culture: Based on the competing values framework*. Upper Saddle River, NJ: Prentice Hall Series in Organizational Development.

Cannon, S. , & Gupta, U. (2020, June). A review of data governance definitions and emerging perspectives [online]. *International Journal of Data Analytics*, 1(2).

Cichy, C. , & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access*, 7, 24634–24648.

Dallemule, L. , & Davenport, T. H. (2017). What's your data strategy. *Harvard Business Review*, 95(3), 112–121.

Floridi, L. , & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A*, 374, 20160360. doi:10.1098/rsta.2016.0360

Gregory, A. (2011). Data governance – Protecting and unleashing the value of your customer data assets. *Journal of Direct, Data and Digital Marketing Practice*, 12(3), 230. doi:10.1057/dddmp.2010.41

Han, J. , Kamber, M. , & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann Publishers, ©2012, Books24x7.

Knoster, T. , Villa, R. , & Thousand, J. (2000). A framework for thinking about systems change. In R. Villa & J. Thousands (Eds.), *Restructuring for caring and effective education: Piecing the puzzle together* (2nd ed.). Baltimore, MD: Paul H. Brookes Publishing Company.

Laranjeiro, N. , Soydemir, S. N. , & Bernardino, J. (2015, November). A survey on data quality: Classifying poor data. In *2015 IEEE 21st Pacific rim international symposium on dependable computing (PRDC)*, Zhangjiajie (pp. 179–188).

Loshin, D. (2001). *Enterprise knowledge management: The data quality approach*. Morgan Kaufmann Publishers.

Mason, H. , Loukides, M. , & Patil, D. J. (2018). *Ethics and data science*. O'Reilly Media. Retrieved from <https://www.safaribooksonline.com/library/view/ethics-and-data/9781492043898/>

Mazzei, M. J. (2017, January 5). Big data dreams: A framework for corporate

strategy. *Business Horizons*, 60(3), 405–414.

Quinn, R. E. , & Rohrbaugh, J. (1981). A competing values approach to organizational effectiveness. *Public Productivity Review*, 5(2), 122–140.

Redman, T. C. (2001). *Data quality: The field guide*. Digital Press.

Schein, E. H. (2010). *Organizational culture and leadership*. The Jossey-Bass Business & Management (Vol. 2). San Francisco, CA: Jossey-Bass.

Sebastian-Coleman, L. (2012). *Measuring data quality for ongoing improvement: A data quality assessment framework* (1st ed.). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Tekiner, F. , & Keane, J. (2013). Big data framework. In *2013 IEEE international conference on systems, man, and cybernetics*, 13–16 October. doi:10.1109/SMC.2013.258

van Es, K. & Schäfer, M. T. (2017). (Eds.), *The datafied society: Studying culture through data*. Amsterdam: University Press.

Wolff, A. , Gooch, D. , Cavero Montaner, J. J. , Rashid, U. , & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics*, 12(3), 9–26.

Index

- Accessibility, [90](#)
- Action plan, [41](#)
- Adaptability, [28](#)
- Adhocracy, [29-30](#)
- Align data strategy, [79-81](#)

- Big data management frameworks
 - analytics, [118-119](#)
 - data sources, [118](#)
 - decision-making, [119](#)
- Business-driven data strategy, [77-79](#)
- Business strategy, [79-81](#)

- Capability maturity model (CMM), [157](#)
- Change management, [15](#), [219](#)
 - data governance, [35](#)
 - definition, [34](#)
- Clan culture, [29](#)
- Common language, [62-63](#)
- Communications, [13](#), [43](#), [91-92](#)
 - agreement, [54-55](#)
 - common language, [62-63](#)
 - context, [50-52](#)
 - definitions, [49-50](#)
 - describing data, [58-60](#)
 - documentation, [55-56](#)

- emotions, [57-58](#)
- foundational work, [48-49](#)
- inform vs. persuade, [65-67](#)
- language, [58](#)
- oral vs. written, [64-65](#)
- people, [61-62](#)
- politics, [56-57](#)
- push vs. pull, [63-64](#)
- repeat vs. reiterate, [67-68](#)
- standards, [52-54](#)

Culture, [14-15](#), [189-191](#), [219](#)

Data collection, [135-137](#)

Data control, [8](#)

Data-driven culture, [41-44](#)

Data ethics, [133-135](#)

Data governance

- big data management frameworks, [117-121](#)
- case studies, [167-185](#), [187-216](#)
- change management, [15](#)
- changes, [11](#)
- chief data officer, [106-108](#)
- cloud computing framework, [116-117](#)
- communications, [13](#)
- culture, [14-15](#)
- data control, [8](#)
- data-related problems, [8-9](#)
- definition, [3-7](#), [218](#)
- elements, [11-13](#)
- frameworks, [101-122](#)
- hard skills, [9](#)
- institutional data, [110-113](#)
- investment, [2-3](#)
- mainstream resources, [2](#)
- organizational placement, [108-110](#)
- research, [1](#), [2](#), [110-113](#)

- responsibility, [15-17](#)
- soft skills, [9](#)
- structure, [103-106](#)
- Data Governance Office (DGO), [200-207](#)
- Data literacy, [129-132](#)
- Data management capability model (DCAM), [158-161](#)
- Data management maturity (DMM), [157-158](#)
- Data quality, [91](#)
 - accuracy, [125](#)
 - assessment frameworks, [127-128](#)
 - completeness, [125](#)
 - consistency, [125](#)
 - data cleansing, [126](#)
 - high quality, [126-127](#)
 - timeliness, [125](#)
 - uniqueness, [125](#)
- Data-related problems, [8-9](#)
- Data transformation, [192](#)
- Defensive strategy, [85-88](#)
- Describing data, [58-60](#)
- Disengaged C-suite, [95](#)
- Documentation, [55-56](#)

- Electronic health record (EHR), [189](#)
- Emotions, [57-58](#)
- Enterprise data warehouse (EDW), [191](#), [193-194](#)
- Ethics, [133-135](#)
 - analytics, [140-141](#)
 - data collection, [135-137](#)
 - data sharing, [137-140](#)
- European General Data Protection Regulation (GDPR), [136](#)
- External relationships, [27](#)

- Facilities Inventory Classification Manual, [50](#)
- Foundational work, communication, [48-49](#)

Hard skills, [9](#)

Hierarchy-oriented culture, [30](#)

HIMSS Analytics Adoption Model for Analytics Maturity, [199-200](#)

HIPPO management, [56](#)

Incentives, [40](#)

Insights, [148-149](#)

Integration, [146-147](#)

Integrity, [147-148](#)

Internal maintenance, [27](#)

IT-driven data strategy, [77-79](#)

Key performance indicator (KPI), [80-81](#)

Known quality, [126](#)

Language, [58](#)

Literacy, [129-132](#)

Market-oriented culture, [30](#)

Maturity models

- capability maturity model (CMM), [157](#)

- caution, [152-154](#)

- comparative, [151-152](#)

- data assessment, [145-146](#)

- data insights, [148-149](#)

- data integration, [146-147](#)

- data integrity, [147-148](#)

- data management capability model (DCAM), [158-161](#)

- data management maturity (DMM), [157-158](#)

- definition, [156](#)

- descriptive, [149-150](#)

- generic steps, [144-145](#)

- IBM maturity model, [162-164](#)

- levels, [154-156](#)
- predictive, [150](#)
- prescriptive, [150](#)

Mechanistic processes, [27](#)

Mission, [28](#)

Offensive strategy, [85-88](#)

O-foods, [82-83](#)

Operational data store (ODS), [194](#)

Organic processes, [27](#)

Organization, [11](#), [188-189](#)

Organizational culture

- adaptability, [28](#)
- adhocracy, [29-30](#)
- clan culture, [29](#)
- cultural models, [26](#)
- hierarchy-oriented, [30](#)
- involvement, [28](#)
- market-oriented, [30](#)
- mission, [28](#)
- multidimensional aspects, [26-27](#)
- person culture, [31-32](#)
- power culture, [30-31](#)
- role culture, [31](#)
- standard organizational behavior, [22](#)
- task culture, [31](#)
- types, [36](#)
- values, [29](#)

Ownership, [90](#)

Person culture, [31](#)

Politics, [56-57](#)

Power culture, [30-31](#)

Responsibilities, [92-93](#)

Role culture, [31](#)

Statistical literacy, [130](#)

Strategy, data

- accessibility, [90](#)

- align data strategy, [79-81](#)

- blame-game, [94](#)

- business-driven data strategy, [77-79](#)

- business strategy, [79-81](#)

- communications, [91-92](#)

- conflict, [94](#)

- cost, [90](#)

- customers, [93-94](#)

- data quality, [91](#)

- disengaged C-suite, [95](#)

- examples, [81-82](#)

- integrity, [91](#)

- IT-driven data strategy, [77-79](#)

- key performance indicator (KPI), [80-81](#)

- offensive vs. defensive strategy, [85-88](#)

- O-foods, [82-83](#)

- organizations, [73](#)

- ownership, [90](#)

- quality, [89](#)

- quantity, [89](#)

- responsibilities, [92-93](#)

- risks, [90-91](#), [95](#)

- roles, [92-93](#)

- source of data, [89](#)

Task culture, [31](#)

Tribal loyalties, [137](#)

Values, [29](#)

Vision, [40](#)