Extracted from:

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

The Pragmatic Bookshelf

Raleigh, North Carolina

The Pragmatic Programmers

Pragmatic exPress

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

Lauren Maffeo

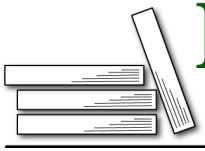
Edited by Brian P. Hogan

Designing Data Governance from the Ground Up

Six Steps to Build a Data-Driven Culture

Lauren Maffeo

The Pragmatic Bookshelf
Raleigh, North Carolina



Pragmatic Bookshelf

For our complete catalog of hands-on, practical, and Pragmatic content for software developers, please visit *https://pragprog.com*.

The team that produced this book includes:

CEO: Dave Rankin
COO: Janet Furlow
Managing Editor: Tammy Coron
Development Editor: Brian P. Hogan
Copy Editor: Karen Galle
Layout: Gilson Graphics
Founders: Andy Hunt and Dave Thomas

For sales, volume licensing, and support, please contact *support@pragprog.com*.

For international rights, please contact *rights@pragprog.com*.

ISBN-13: 978-1-68050-980-9
Encoded using the finest acid-free high-entropy binary digits.
Book version: P1.0—January 2023

Preface

I don't need to tell you how essential data is. If you picked up this book, then you already know that data powers today's most proftable businesses. With companies like Google and Meta boasting data-based business models, it's no surprise that 99 percent of firms said they planned to invest in AI and data in 2021.[1]

While data can bring incredible value, most businesses haven't unlocked it yet. Poor data quality costs businesses millions per year, with some paying close to a quarter of their annual revenue.[2] If that's not enough of a reason to change, consider the cultural impact of bad data.

A 2021 survey of C-suite executives found that while investments in AI and big data keep rising, the number of respondents who say they're data-driven declined. Just one in four respondents said they thought their organization was data-driven, down from 37.8 percent the year before.[3] Those surveyed consistently cited cultural hardships as a far bigger blocker than technical limitations. Employee skills gaps, outdated business processes, resistance to change, and poor choice of tools add up over time. If you've wondered why just 13 percent of machine learning models make it to production, there's no shortage of culprits.

I've seen this firsthand in my own work. I spent several years researching AI techniques as an analyst at Gartner. My job was to advise clients on the latest technical trends that could help them grow their businesses. I quickly realized that even if most of these businesses started using AI tomorrow, their efforts would be futile. Most businesses had such low data maturity that they weren't ready to harness technology like machine learning.

1. https://c6abb8db-514c-4f5b-b5a1-fc710f1e464e.filesusr.com/ugd/e5361a_76709448ddc6490981f0cbea42d51508.pdf
2. https://sloanreview.mit.edu/article/seizing-opportunity-in-data-quality/
3. https://hbr.org/2021/02/why-is-it-so-hard-to-become-a-data-driven-company

When I started working on a technical team, I saw how even organizations that exist to disseminate data are not immune to governance problems. I've seen clients possess millions of unique data points and several centuries' worth of datasets. They also lacked any documentation showing which servers this data lived on, how these servers integrated with each other, and the workflows that shared this data with their users. As a result, their data dissemination processes were not streamlined, took days to complete, and involved several people per release, with no automation involved.

It doesn't have to be this way. Consider the efforts involved in bringing a new product to market. You would write a go-to-market plan with a high-level strategy for how you want the product to help users. You would work with colleagues across sales, marketing, engineering, and customer success to collect intel that will help you build this product. You would write a roadmap showing the key tasks and milestones each stakeholder must meet to bring this product to the right customers at the right time. Then, you would keep governing this product throughout its life cycle.

Your data strategy deserves no less attention than your product strategy. The good news is, a lot of techniques needed to build great products apply to data governance as well. Mastering the basics—finding a framework, selecting data stewards, building your data governance council, and writing a roadmap—helps you take your data projects to the next stage of maturity.

This book is your blueprint to bring data projects past production. By building and executing a plan to use data for decision-making, you'll gain the first six steps you need to build a data governance plan that improves business outcomes and engages colleagues. By putting this into practice, you'll build trust, increase cooperation, and improve efficiency when it comes to using data. My goal is for you to start reading this book on your flight from Los Angeles and land in New York with the steps you need to start building a data governance plan tomorrow.

Is This Book for Me?

I wrote this book for decision-makers in medium-to-large organizations (100+ employees), especially those working in highly regulated industries where data governance is legally required. I believe that all organizations should practice data governance, regardless of size or industry. It's easier to build programs from scratch than to incur the technical debt involved in fixing broken practices. That said, you will gain the most from this book if you already work in an organization that has the staff numbers needed to execute governance. While you don't need an enormous team, you can't do all of this work alone.

I wrote this book for readers who know they need help leveraging their data, but are not sure how to get governance off the ground. It's for readers who hold decision-making power at their organizations, and are tasked with finding solutions for business problems. If the CEO of your accounting firm wants advice on how to take your robotic process automation project past production, or your higher education client's asking how to manage metadata, this book will make the case for you.

That reader definition is intentionally broad. It's true that I wrote this book for senior leaders in business and tech roles to help them build blueprints for data governance. But if a junior data engineer or mid-level manager finishes this book without learning how to harness data in their organizations—and how to engage colleagues in this effort—I don't think I've done my job. In today's organizations, there's no role that data doesn't impact. You might not be an engineer or CTO, but we need you to succeed just the same.

That's especially true if you work in a highly regulated industry like finance, education, or healthcare. While all organizations must use data responsibly, global standards demand that some sectors practice more data discretion. You'll find this book especially helpful if you work in a sector that's controlled by a range of governance rules. Data protections differ widely around the world. If your organization wants to serve a global market, your data governance must meet global standards.

Finally, if you're a student enrolled in a data science or analytics program, I hope this book will help you walk into your next role with the confidence to lead crucial conversations. I once spoke at a data science conference which touched on some of the principles we'll cover in this book. When I asked attendees with Data Science Master's degrees how much instruction they got on data governance, they all said, "None."

They had attended top academic programs and earned data science roles at attractive companies once they graduated. Yet the more I spoke with them, the more I heard how they were expected to do all things data without the knowledge they needed to drive substantial change. These conversations taught me that data science is now too big a burden for one colleague to carry. Without help from the whole organization, today's data scientists will collapse under its weight.

Now that I've told you whom this book is for, it's worth mentioning what's absent. We won't debate what data science is, discuss data training techniques like linear regression, or explain how to do ensemble modeling, because that knowledge exists elsewhere. I wrote this book because you can throw a

proverbial stone online and hit a blog post about data science, but results for data governance searches are scarce. This book fills the void I found when I searched for data governance resources that I could use in my own work. I wrote the book I wished I'd had, based on knowledge I learned the hard way.

I'll reference technical tools and methods where applicable, like Hadoop software utilities and two-factor authentication. The last two chapters are more technical than the first four; they'll cover how to practice data governance once your project is in development and production environments. You'll gain the most value from this book if you're familiar with these concepts. That said, you shouldn't need to hold a traditionally technical role (like data engineer or front-end developer) to learn something from these chapters and this book. If you're a designer, program manager, or quality assurance engineer working on data projects, I hope you'll bring this book's knowledge back to your teams.

What's in This Book?

When you finish this book, you'll have the first six steps you'll need to build a data governance plan that aligns your efforts to clean, collect, and secure the data you'll need to power pipelines. Here's what you'll find in each chapter:

Chapter 1: Find Your Data Framework

How will your organization control data access, implement new regulatory policies, and track activity for data models? If you lack answers to any of these questions, it's time for a data governance framework. This chapter will show you how to use one, and write a mission statement for data use to guide your work through the rest of this book.

Chapter 2: Select Data Stewards

Data governance isn't one sole person or team's job. This chapter will share how to find the best data stewards in your business to own specific types of data and metadata so that data governance is truly a team effort.

Chapter 3: Build Your Data Governance Council

Data stewards can't collaborate if they stay in silos. Chapter 3 shares how to create and manage a data governance council that keeps colleagues engaged for the long haul.

Chapter 4: Write Your Data Roadmap

Roadmaps build business-wide consensus for how to use data, approve tools, prioritize projects, and more. This chapter will show you how to write a roadmap for your data governance work.

Chapter 5: Practice Governance-Driven Development

Data governance is futile if it doesn't improve your development work. Using a case study from Netflix, you will learn how to embed data governance principles into your data projects.

Chapter 6: Monitor Data in Production

Data governance doesn't stop once your models reach production. This chapter shares the steps your stewards must take to keep data's quality and security aligned with your governance standards.

How Will This Book Help Me Reach My Goals?

This book shows you how to build a foundation for collecting, securing, and managing your organization's data. You can't progress your data projects without doing some essential groundwork first.

Establishing data governance that's strategic, follows a framework, engages stewards, and uses a roadmap is hard work. It takes time, resources, and company-wide buy-in. This book will show you how to do it right the first time and set yourself up for long-term success.