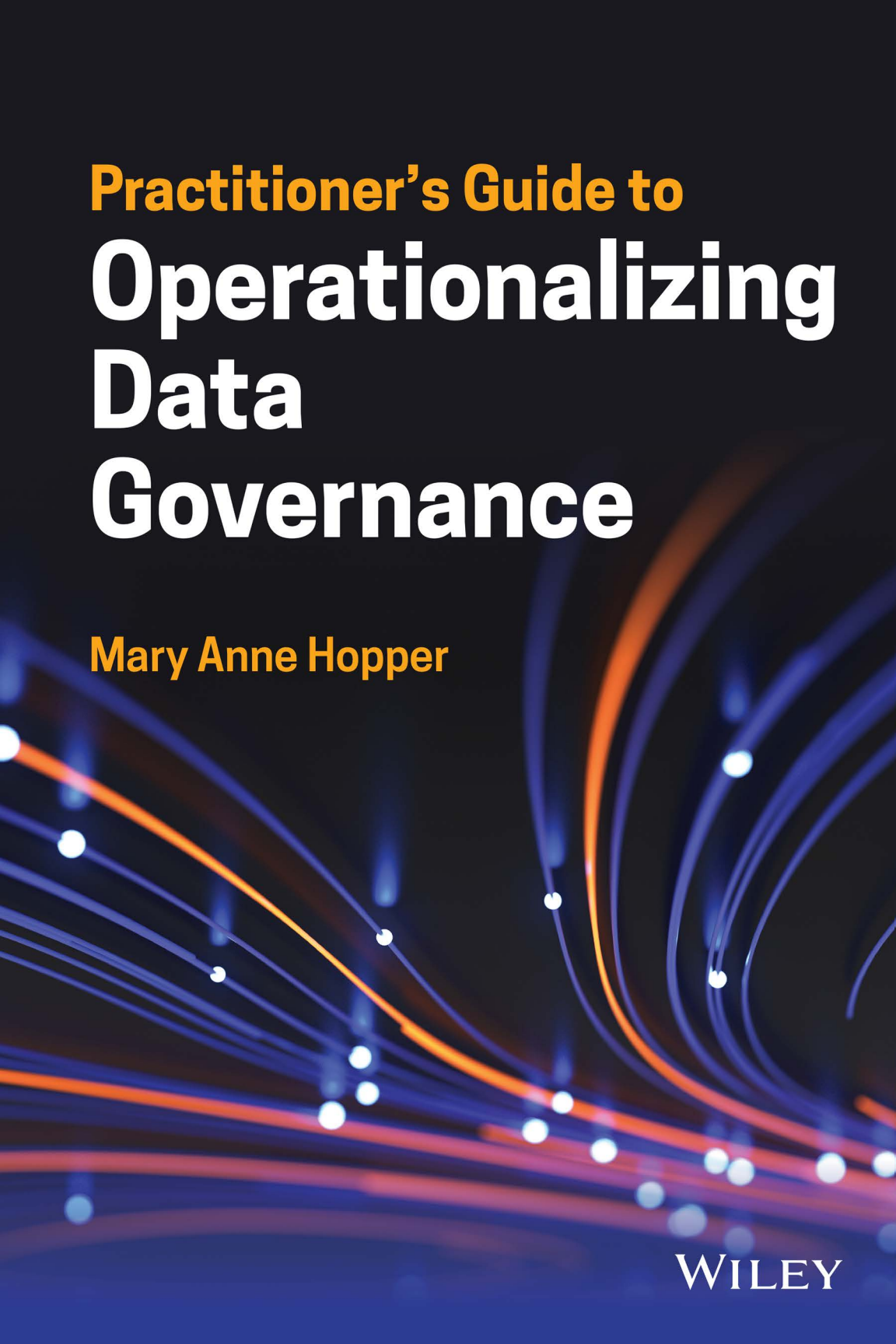
Practitioner's Guide to Operationalizing Data Governance

Mary Anne Hopper



WILEY

Practitioner's Guide to Operationalizing Data Governance

Wiley and SAS Business Series

The Wiley and SAS Business Series presents books that help senior level managers with their critical management decisions.

Titles in the Wiley and SAS Business Series include:

The Analytic Hospitality Executive: Implementing Data Analytics in Hotels and Casinos by Kelly A. McGuire

Analytics: The Agile Way by Phil Simon

The Analytics Lifecycle Toolkit: A Practical Guide for an Effective Analytics Capability by Gregory S. Nelson

Anti-Money Laundering Transaction Monitoring Systems Implementation: Finding Anomalies by Derek Chau and Maarten van Dijck Nemcsik

Artificial Intelligence for Marketing: Practical Applications by Jim Sterne

Business Analytics for Managers: Taking Business Intelligence Beyond Reporting (*Second Edition)* by Gert H. N. Laursen and Jesper Thorlund

Business Forecasting: The Emerging Role of Artificial Intelligence and Machine Learning by Michael Gilliland, Len Tashman, and Udo Sglavo

The Cloud-Based Demand-Driven Supply Chain by Vinit Sharma

Consumption-Based Forecasting and Planning: Predicting Changing Demand Patterns in the New Digital Economy by Charles W. Chase

Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS by Bart Baesen, Daniel Roesch, and Harald Scheule

Demand-Driven Inventory Optimization and Replenishment: Creating a More Efficient Supply Chain (Second Edition) by Robert A. Davis

Economic Modeling in the Post Great Recession Era: Incomplete Data, Imperfect Markets by John Silvia, Azhar Iqbal, and Sarah Watt House

Enhance Oil & Gas Exploration with Data-Driven Geophysical and Petrophysical Models by Keith Holdaway and Duncan Irving

Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection by Bart Baesens, Veronique Van Vlasselaer, and Wouter Verbeke

Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards (Second Edition) by Naeem Siddiqi

JMP Connections: The Art of Utilizing Connections in Your Data by John Wubbel

Leaders and Innovators: How Data-Driven Organizations Are Winning with Analytics by Tho H. Nguyen

On-Camera Coach: Tools and Techniques for Business Professionals in a Video-Driven World by Karin Reed

Next Generation Demand Management: People, Process, Analytics, and Technology by Charles W. Chase

A Practical Guide to Analytics for Governments: Using Big Data for Good by Marie Lowman

Profit from Your Forecasting Software: A Best Practice Guide for Sales Forecasters by Paul Goodwin

Project Finance for Business Development by John E. Triantis

Smart Cities, Smart Future: Showcasing Tomorrow by Mike Barlow and Cornelia Levy-Bencheton

Statistical Thinking: Improving Business Performance (Third Edition) by Roger W. Hoerl and Ronald D. Snee

Strategies in Biomedical Data Science: Driving Force for Innovation by Jay Etchings

Style and Statistics: The Art of Retail Analytics by Brittany Bullard

Text as Data: Computational Methods of Understanding Written Expression Using SAS by Barry deVille and Gurpreet Singh Bawa

Transforming Healthcare Analytics: The Quest for Healthy Intelligence by Michael N. Lewis and Tho H. Nguyen

Visual Six Sigma: Making Data Analysis Lean (Second Edition) by Ian Cox, Marie A. Gaudard, and Mia L. Stephens

Warranty Fraud Management: Reducing Fraud and Other Excess Costs in Warranty and Service Operations by Matti Kurvinen, Ilkka Töyrylä, and D. N. Prabhakar Murthy

For more information on any of the above titles, please visit www .wiley.com.

Practitioner's Guide to Operationalizing Data Governance

Mary Anne Hopper

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Names: Hopper, Mary Anne, author.
Title: Practitioner's guide to operationalizing data governance / Mary Anne Hopper.
Description: Hoboken, New Jersey : Wiley, [2023] | Series: Wiley and SAS business series | Includes index.
Identifiers: LCCN 2023001522 (print) | LCCN 2023001523 (ebook) | ISBN 9781119851424 (cloth) | ISBN 9781119851455 (adobe pdf) | ISBN 9781119851431 (epub)
Subjects: LCSH: Database management. | Management information systems—Management. | Data integrity.
Classification: LCC QA76.9.D3 H6564 2023 (print) | LCC QA76.9.D3 (ebook) | DDC 005.75/65—dc23/eng/20230201
LC record available at https://lccn.loc.gov/2023001522
LC ebook record available at https://lccn.loc.gov/2023001523

Cover Design: Wiley
Cover Image: © shulz/Getty Images

For the MAC team (Faramarz, Matthias, Matt, Katie, Noah, and Liz) for all your support

and

For Bill for never letting me throw the term "best practice" out there without explaining "the why".

Contents

Acknowledgments

A methodology does not create itself overnight. It takes time, encouragement, and dedication from a lot of people. The methodology I have laid out for you is from the work and guidance of my colleagues at the SAS Institute: Faramarz Abedini, Matthias Gruber, Robert Stone, Matt Benson, Katie Lorenze, Noah Pearce, and Liz Baker. Some of you I have worked with for many years, some not quite as long. Either way, your fingerprints are embedded in the content of these pages. For that, I thank you!

CHAPTER 1

Introduction

INTENDED AUDIENCE

As long as the practice of Data Governance has been around, the concept continues to lack sustainable adoption in many organizations. My main objective with this book is to share my experience and help you and your organization on your journey, no matter where in that journey you are.

My best guess is that you are looking at this book as a guide for one of the following reasons:

- Your organization is thinking about Data Governance.
- You have been tasked with Data Governance.
- You need to get your Data Governance program back on track.
- You have acquired a tool and want to get the most value from your investment.
- You continue to have the same data quality issues over and over.
- You attended a conference and learned about Data Governance and think it is something you need.

The content in this book is meant for a large audience because Data Governance impacts the entire organization. Whether a senior leader or an individual contributor, you may be asked to participate at some level in Data Governance, actively or passively.

This book guides you through practical steps in applying Data Governance concepts to solve business problems by adopting a disciplined approach to Data Management methods. The chapters cover prioritization, alignment of Data Governance and Data Management, organizational structures, defining roles and responsibilities, communications, measurements, operations, implementation, and policies. All of the examples presented are not conceptual; they are real-world customer examples that can be applied to your specific organization.

EXPERIENCE

You most likely have an interest in not just Data Governance, but in data itself. Do you remember your "Aha" moment that turned you into a data junkie? I remember mine clearly. In the early 1990s, I

worked for a small naval architectural firm. The focus of the firm was primarily custom high-end racing sailboat designs, including the America's Cup. One day my boss brought in a floppy disk and asked me to take a look at what was on it. Apparently, we had a client who thought his brand-new boat was slow. The disk contained the data dump from the boat's instruments. There were fields like time of day, heading, wind velocity, and boat speed. I was able to parse the data and essentially recreate the races with the available data points. What I learned was that the boat tacked nine or ten times on the first leg of each race. I know not all of you are expert sailboat racers but take my word for it; tacking that many times on any leg of a race in a big boat is slow. What did that mean for my boss? He was able to have a different conversation with our client. We were no longer defending boat design or building materials but instead talking about racing tactics and offering suggestions for improvements there first.

That day changed my view of the power of data and from that point forward I chose classes and career roles that were focused on data. Initially, I focused on database development and support and then transitioned into data warehouse development. On the IT side, I managed the development of platforms to support finance and treasury processes as well as the re-platforming of a home-grown loan servicing system. That experience enlightened me to the need for data quality processes and the understanding of data lineage and documented business rules. There came a time when I transitioned into project management, product ownership, and finally consulting. The consulting role is what has helped me most in hearing customer challenges and helping them solve those problems by instilling discipline in Data Management processes.

Over the years, I have worked with hundreds of clients across all industry verticals to help them establish that discipline in Data Management practices. In other words, helping them to establish Data Governance programs that align with their individual organization's business objectives while also considering their maturity, culture, and appetite for Data Governance.

This book is not only a reflection of a tested and proven methodology but also my experiences in what works and what doesn't work, things to not get hung up on, and where best to focus efforts. Some of

the chapters are shorter than others but I still believe the topics are important enough to cover. My hope is that this book helps you and your organization in your own Data Governance journey.

COMMON CHALLENGE THEMES

Most of what I've heard over the years can be broken down into a set of common themes. One of the best ways to talk about those themes is to share with you what I've heard my clients say. Every quote is directly from a customer. If any of these quotes resonate with you, then formalizing Data Governance can help. You will see these themes again in future chapters.

Metadata

Metadata is the practice of gathering, storing, and provisioning information about data assets. As important as it is to collect and maintain, it is a practice that does not formally exist in most organizations. Most of my customers might not necessarily use the term metadata, but the concept is top of mind for them. There is a desire to have common terms defined and have a single repository to maintain information about those terms. Because there is no formal metadata process or repository, users spend a lot of their time trying to understand data on their own or relying on others to interpret meaning for them. Another byproduct from the lack of metadata process is that users complain of not knowing what data is available to them. Always keep in mind that metadata is a precursor to data quality; I will write more about that topic in later chapters.

Here is what clients have said:

- "we need Rosetta Stone for our data"
- "metadata is so important and it doesn't exist"
- "the most time-consuming part is to find what you're looking for"
- "would be nice to follow the trail"
- "can't get to confident decisions without common definitions"

- "a little bit of detective work and a little bit of knowledge"
- "this is what I mean when I say 'this'"
- "we haven't the foggiest idea of what the denominator is"
- "you get the data and it's not what you meant"
- "some people just want to call it something different"

Access to Data

Oftentimes, there are very few people with the "know-how" and the tools to access data. Users who do have direct access feel they must navigate a labyrinth to get to the data they need. That labyrinth includes multiple reports, accessing tables, or calling people who have knowledge of data structures. Because of this, users find it easier to maintain their own datasets instead of accessing a common repository. In most organizations, users are anxious to have access to tools to make it easier to use data.

Here is what clients have said:

- "we got to know what the hell we got"
- "our issue isn't so much storage, it's access"
- "quit parking data on some machine"
- "a whole lot of horsepower to pull data out of that system"
- "you have to have your DNA tested before you get access to it"
- "not knowing something exists is a greater liability than not using what is available"
- "a lot of what we're doing seems so hard"
- "information does not seem readily available"
- "manual data exercise to put it together"
- "we have so much information out there in so many places"
- "Excel becomes the big workhorse"
- "we've created a process to deal with lack of access to information"
- "want to hire an analyst, not a SQL person"
- "high-priced analyst just getting data for people"

Trust in Data

Users want the ability to make solid decisions on trusted data that is deemed a definitive source of truth. However, users feel there is a lack of consistency across data sources. Some of the reasons for this could be related to data latency, poor data collection practices, a lack of data understanding (e.g., data acceptance, service level agreements, data remediation, and data profiling), or different groups creating and maintaining their own copies of data. This results in users feeling they spend a significant amount of time validating or defending the data they do use.

Here is what clients have said:

- "depending on which query you run you get a different answer"
- "can't create individual sources of truth"
- "the place we pull the data from doesn't balance to itself"
- "we don't know how reliable the data is"
- "you trust the data until you know it's not right"
- "if you can't fix the problem you work around it"
- "how do we know what an error looks like?"

Data Integration

Data integration consists of processes for moving and combining data that reside in multiple locations and providing a unified view of the data. In many environments, users who need access to integrated data are essentially required to pull several reports or datasets and then integrate on their desktop using MS Access or Excel. There may also be a lack of formal processes or tool usage across divisions and even in IT. More often than not, this results in differing business rules that are applied to data, which turns into discrepancies in the data results.

Here is what clients have said:

- "I'm living in spreadsheet hell"
- "really no linking it all together"

- "right now, it's fragmented"
- "all our stuff doesn't talk to each other"
- "being able to stitch data together is what we need"
- "our systems have never been organized to allow us to answer questions"
- "almost every prototype that we did last fiscal year had to do with the difficulty of pulling data from multiple datasets"
- "we have a lot of questions, we have a lot of data, but we can't pull it out easily"

Data Ownership

Users do not know who to contact when there are data questions. There is a desire to have a named data owner for the various domains who can answer questions, address issues, and help users understand data usage guidelines for given datasets.

Here is what clients have said:

- "you're stepping on toes every time you go in there"
- "everybody wants to control their own fate"
- "that's our data so we should be able to keep up with it"
- "lack of accountability for data responsibility"
- "we don't really know who does that"

Reporting/Analytics

Users are becoming more data aware. Although some users only require operational reports, there is a growing curiosity and desire for more advanced analytic capabilities. This makes the reporting and analytic platform (e.g., data warehouse, data mart, date lake, etc.) being part of the overall strategic plan more important than ever. Most users feel it takes up a lot of their time to get reports and like the concept of a single point of entry for all of their reports as opposed to reports within the various applications that they are forced to self-integrate.

Here is what clients have said:

- "how much of that data is relevant to the next level of the department?"
- "I don't think people realize what we could do [if our data were integrated]"
- "very myopic view of the data"
- "[we need to be able to provide] reliable, repeatable answers to questions"
- "the most time-consuming part is to find what you're looking for"
- "would like to have a dashboard to share accurate information"
- "it's more art than science"
- "information is perceived as ad-hoc"
- "used to managing without information"
- "too much reliance on old data to make current decisions"

Data Architecture

While most organizations have an architecture practice in place, the teams often lack authority because they do not have a formally defined charter. There is no formal data strategy to help set the team's direction and enable it to define standards and guidelines for identifying, provisioning, storing, and integrating data. With newly formed teams especially, the focus is on new applications instead of the entire enterprise data landscape that has been growing for years with no formal practices in place.

Here is what clients have said:

- "there have been so many architecture hands over the years"
- "we are duplicating a lot of information"
- "[there is] no logic in how we approach managing data"
- "data should be accessible regardless of where its source is"
- "it's extremely laborious"
- "tendency to work like we're all artisans here"