

Brought to you by:



# Data Catalogs

**for  
dummies®**  
A Wiley Brand



**Collibra  
Special Edition**

Learn what a  
data catalog is

Easily find and  
understand your data

Deliver trusted  
data insights

**Bineesh Babu**

**Ben Moser**

**Chandra Papudesu**

**Luis Romero**

# Table of Contents

INTRODUCTION .....	1
About the Book.....	1
Foolish Assumptions.....	2
Icons Used In This Book.....	2
Beyond the Book.....	2
CHAPTER 1: <b>Overcoming Today's Data Challenges</b> .....	3
Identifying Common Data Management Challenges.....	3
Dealing with increasing volume, variety, and velocity.....	4
Meeting ever-more-stringent privacy rules .....	5
Spending valuable time on bad data.....	5
Struggling with data pipeline management .....	6
Data Catalogs to the Rescue .....	7
Understanding the purpose of a data catalog .....	8
Why are there so many data catalogs and providers? .....	9
Why you need an enterprise data catalog.....	9
Considering the users .....	10
CHAPTER 2: <b>Determining the Right Data Catalog for Your Organization</b> .....	11
Setting Your Goals.....	12
Identifying the right metrics .....	13
Identifying the Three Stages of a Data Catalog .....	14
Crawl: Ensuring you have the basics covered .....	14
Walk: Taking it to the next level .....	15
Run: Making your data catalog work for the enterprise .....	15
CHAPTER 3: <b>Maximizing Your Data Catalog's Value</b> .....	17
Setting Priorities .....	17
Leveraging Your Data Catalog for Data Governance and Privacy Initiatives .....	19
Curating Your Data Collections.....	20
Ensuring Trustworthy Data .....	21
Creating an Integrated Experience .....	22

CHAPTER 4: **Powering a Data Intelligence Platform** ..... 23

    Creating a Data Intelligence Strategy..... 24

    Activating Your Metadata Graph ..... 25

        Enriching your metadata graph with policies..... 26

        Operationalizing your data catalog with workflows ..... 27

CHAPTER 5: **Evaluating Your Data Catalog’s Success**..... 29

    Understanding Why KPIs Are Important ..... 30

    Categorizing the Types of KPIs ..... 30

        Enablement metrics: Deployment and education..... 31

        Adoption metrics: Usage..... 32

        Business value metrics: Quantifiable outcomes ..... 35

    Following Three Essential Strategies..... 36

CHAPTER 6: **Seven Steps to Data Intelligence** ..... 37

    Support the Entire Enterprise..... 37

    Connect to All Your Data ..... 38

    Enable Automation at Scale ..... 38

    Gain Visibility with Data Lineage ..... 38

    Drive Trust with Collaboration..... 39

    Embed Data Governance and Privacy ..... 39

    Establish a Data Marketplace ..... 40

# Introduction

**A**re you drowning in data? It seems like most organizations are, in this data-driven era.

Organizations know they need to become more data-driven to maintain profitability, support innovation, and grow the business. The most successful organizations in the world — from governments to global corporations — have made data a strategic asset. In fact, a January 2021 report from Forrester, “Chief Data Officers: Evolve Your Teams To Accelerate Impact From Data Insights,” states that 63 percent of organizations have appointed a Chief Data Officer.

But there are many headwinds in the form of challenges and risks that stand in their way. Chief among these challenges is the scale and complexity of the data ecosystem. As more sources of data come online, the problem only becomes worse. Then there are the threats — both internal and external — to security and reputation.

What’s the solution? For many organizations, it’s implementing a *data catalog*. Data catalogs pull everything together from across your entire digital environment and help you make sense of it, at both the macro and micro levels.

If that sounds interesting, and perhaps a good fit for your organization, you’ve come to the right place!

## About the Book

*Data Catalogs For Dummies* explains why organizations are increasingly turning to data catalogs, and how they can address today’s data challenges. You’ll learn about:

- » Taking the first steps to control your data landscape
- » Identifying the right data catalog for your organization
- » Maximizing the value of your data catalog
- » Leveraging your data catalog for data governance and privacy

- » Turning your data catalog into a full-featured data intelligence platform
- » Defining and measuring success metrics and outcomes

## Foolish Assumptions

We've tried to make this book as accessible and readable as possible for everyone. We've designed the content of this book for anyone that wants to learn how the right data catalog can help them discover, understand, and extract insight across their organization's data.

## Icons Used In This Book

To help understand data catalogs a little better, we use these icons to point out information we want you to pay attention to:



REMEMBER

We use this icon when there's something important you should commit to memory before you progress on with the chapter.



TIP

You find these icons dotted throughout the book giving you helpful suggestions and bite-sized nuggets of useful information.



WARNING

We don't just know what to do. We also know what not to do. Look out for these icons to help you avoid some common pitfalls.



TECHNICAL  
STUFF

This is the jargon buster icon. When we have to get technical, we'll give a little plain English explanation of what we mean.

## Beyond the Book

Want to find out how you can take the next step along your data intelligence journey? You can visit [www.collibra.com/data-catalog](http://www.collibra.com/data-catalog) to learn more and sign up for a tour of Collibra Data Catalog.

## IN THIS CHAPTER

- » Understanding today's data-driven imperative
- » Analyzing the challenges that volume, variety, and velocity present
- » Identifying the purpose and utility of a data catalog
- » Understanding why legacy approaches can't solve your current data management challenges

# Chapter 1

## Overcoming Today's Data Challenges

**T**he status quo of the last decade — in other words, “the way we’ve always done it” — isn’t serving today’s business data needs very well, and many businesses are feeling the pain. In this chapter we review some of the challenges businesses are facing, and then take a first look at how data catalogs can help.

### Identifying Common Data Management Challenges

Today’s organizations are beset on all sides with more data, more demands to derive value from it, and less certainty about how to more efficiently protect and access this data. This section takes a look at some of these challenges to quantify what you’re up against.