opentext™

Managing data bias, privacy, and drift:

The importance of holistic data management

Contents

Introduction

Organizations today are dealing with data and application sprawl; in fact, in just over a decade, the volume and velocity of data, compounding with the arrival of GenAI, is exponential. The world generates 402.74 million terabytes of data per day, and by 2025, the amount of data generated daily will reach 181 zetabytes globally.[1]

Meanwhile, data breaches increase in frequency and impact, privacy regulations are becoming more stringent, and customers demand transparency on how their data is being used and protected. Modern data-fueled organizations need higher levels of visibility, control, and security over their data to remain competitive, efficient, and agile.

Despite evolving challenges, leading organizations should seize opportunities and act now to not only survive, but also thrive in this environment.

Forward-thinking organizations strategically approach data discovery and classification to understand where their data is, the value and risk this data represents, and proactive measures that help to contain costs, detect and protect sensitive data, comply with regulatory and privacy policies, and more.

The following are five strategic best practices and actionable reasons to deploy a holistic data discovery and management program. Armed with these strategic insights, organizations can overcome the challenges of data sprawl and position themselves for accelerated growth in a rapidly changing data landscape.

Critical questions

- Where is all your data?
- What data is considered sensitive?
- Who has access to the data and what level of access?
- When is data being shared?
- How is the data managed?

[1] Exploding Topics, Amount of Data Created Daily, 2024.

Strategy #1

Identify and categorize sensitive data across diverse environments to ensure comprehensive oversight.

Today, data is everywhere—on premises, in the cloud, in archive, with third parties, and in use. The arrival of generative AI is also contributing new data to this multi-dimensional data sprawl.

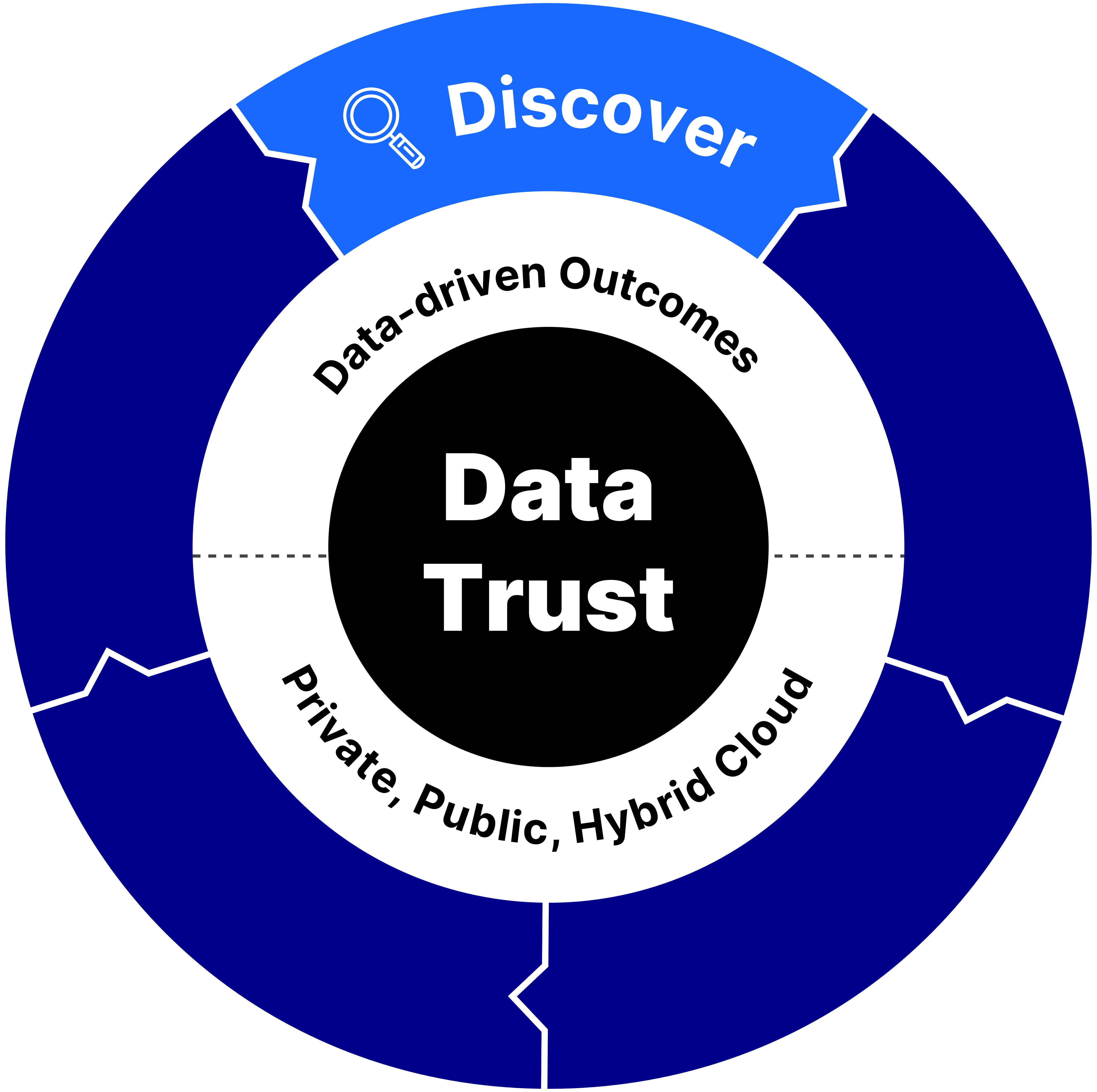
Pertinent questions arise: Where is all this data located? Is it still being used? What about data not being used?

It all starts with data discovery to find and classify data.

Accuracy is crucial to create meaningful insights to help drive the business forward because not all data is equal. It's vital to deploy a data discovery and classification program that identifies all the data from disparate environments.

However, data discovery capabilities aren't enough on their own.

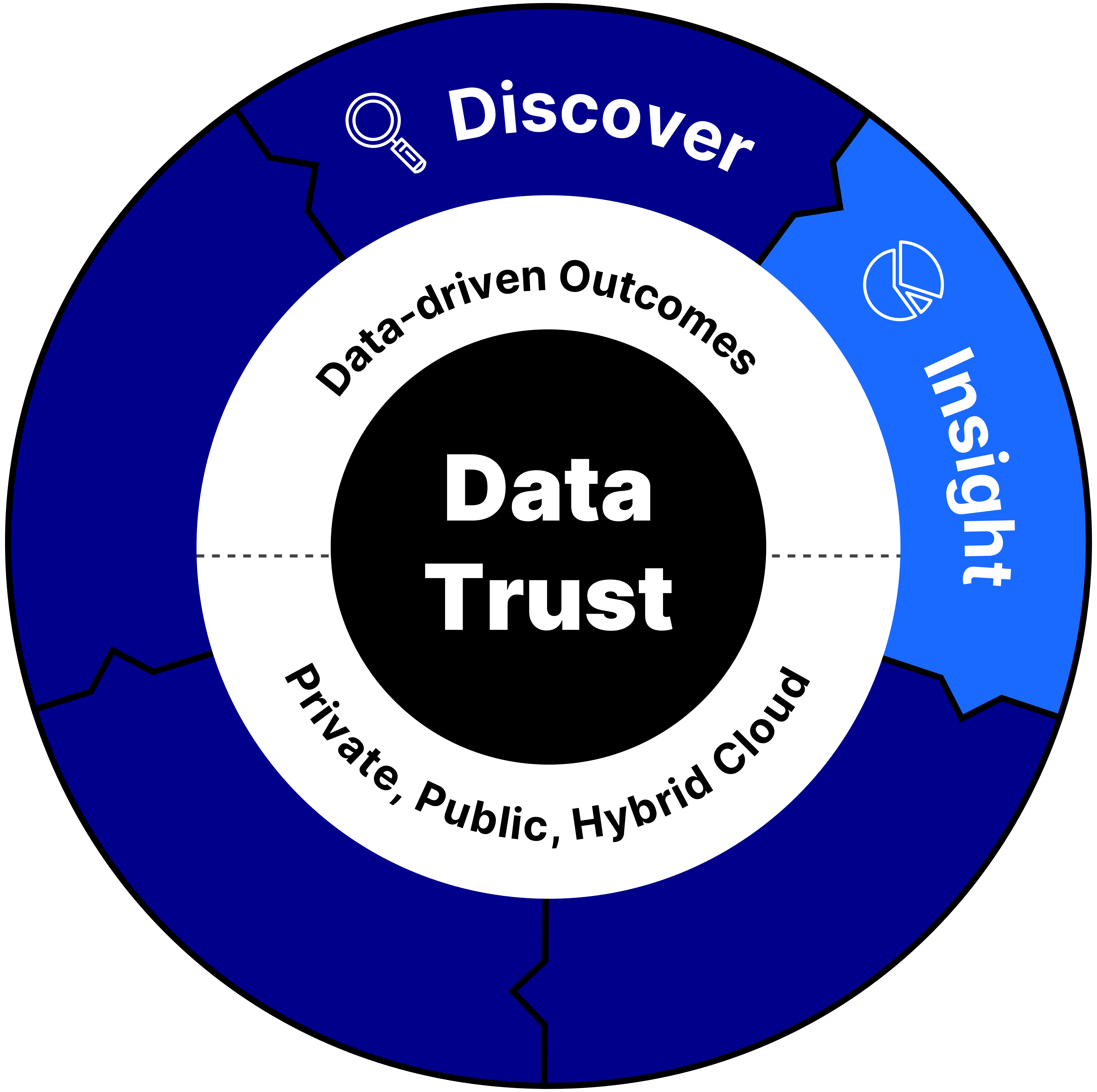
Organizations need to first find where the sensitive data, like personal identifiable information (PII), is concentrated, then accurately categorize or safely move and secure the data. Data discovery and curation need to work together. Successful organizations champion a holistic program that performs data discovery, classification, and more.



Strategy #2

Collect data with a clear purpose and make sure user consent aligns with ethical standards and legal requirements, fostering trust and transparency.

Customers trust you with one of their most important assets: their data. They want to know how the data is being collected, used, and safeguarded. Meanwhile, organizations face pressure from governments and legal frameworks to responsibly govern and process sensitive PII. Organizations are making the effort to comply with data regulation. However, a lot of organizations focus on structured data, while 80–90% of data is unstructured, according to "The untapped value of unstructured data."[2]

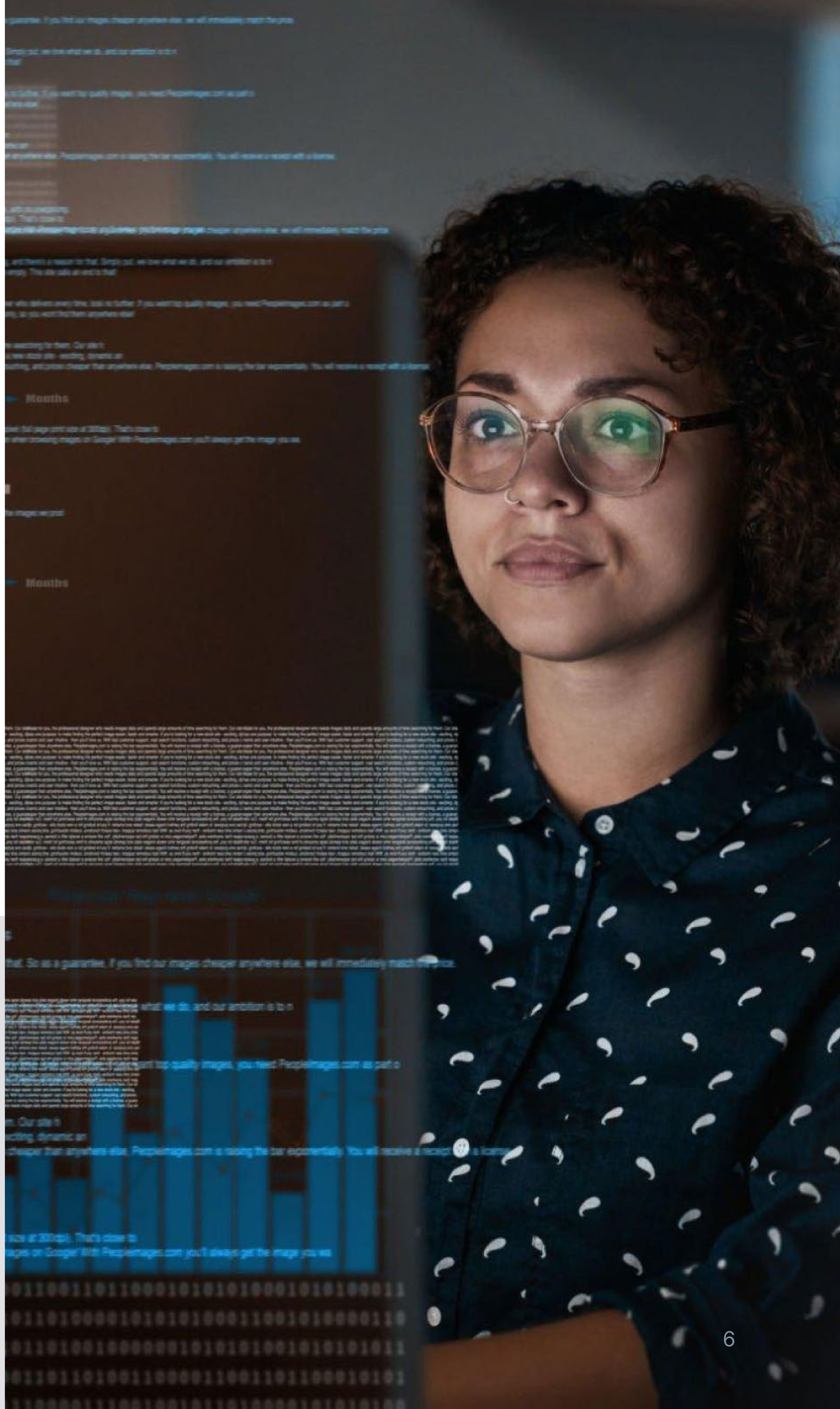


[2] IDC, Untapped Value: What Every Executive Needs to Know About Unstructured Data, 2023.

Here’s what to do

- Deploy a strategic data collection and preservation program to collect the data (including unstructured) from disparate sources, gain early insight, and ensure secure workflows.
- Defensibly collect and preserve data at scale for litigation and investigations, but in a way that is responsible and strategic for your business.
- Incorporate format-preserving data protection. This includes the conversation of sensitive data elements into non-sensitive replacement values that are the same length and format as the original data elements.

Discovering, categorizing, responsibly collecting, and safeguarding data are all critical and interconnected strategies successful organizations embrace.



Data tokenization, masking, and encryption

The terms “tokenization,” “data masking,” and “encryption” are all often used generically to mean any form of format-preserving data protection. Learn more.

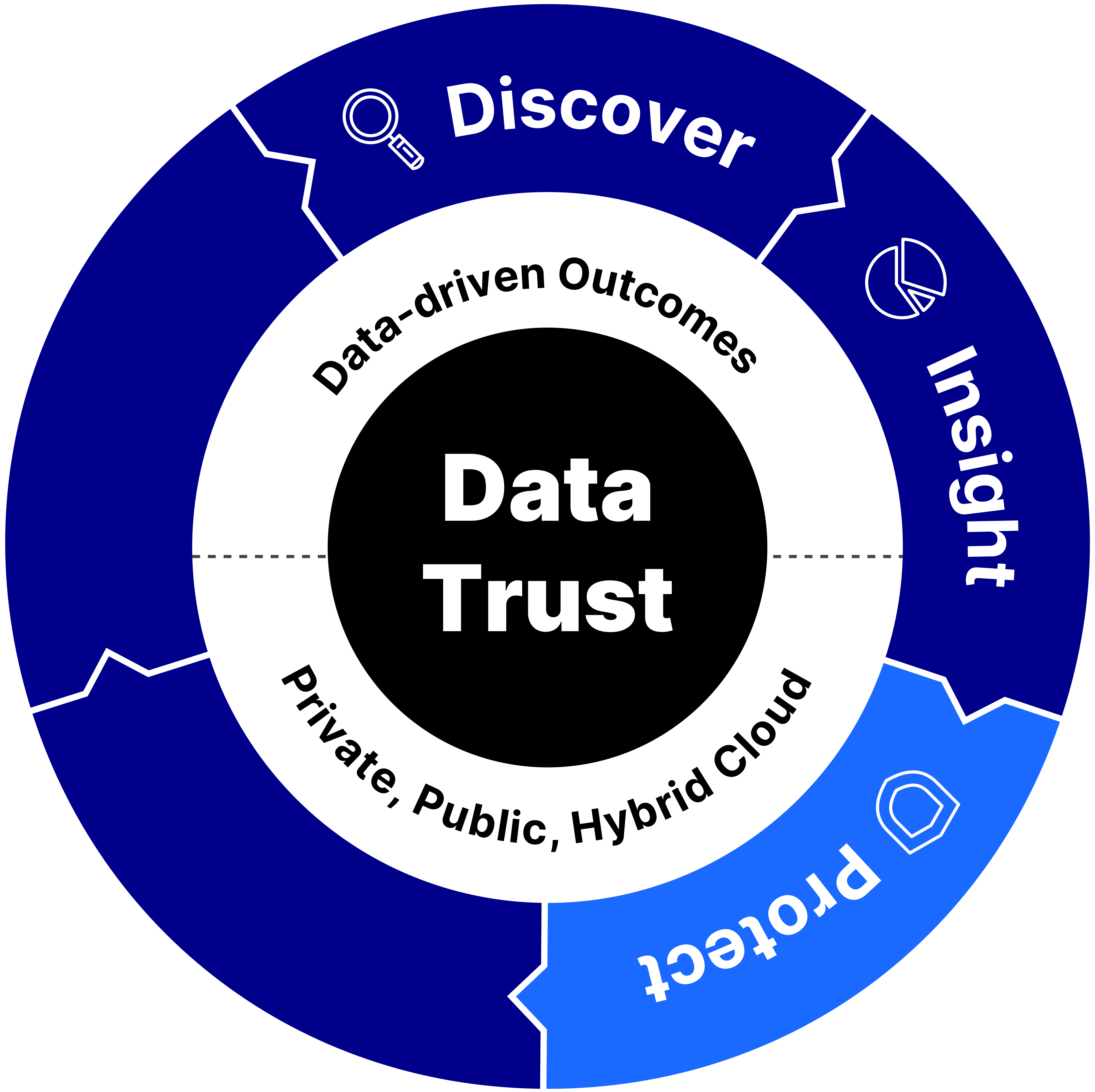
Strategy #3

Balance safeguarding personal data with business utility.

Cybersecurity—and in particular, data protection—is top of mind for the C-suite, the board, and customers too. You discovered data, categorized it, and collected the data ethically and for legal reasons you must determine what to protect and what data to delete.

Managing redundant, obsolete, trivial (ROT) data: unnecessary, duplicated, and outdated information.

As mentioned, not all data is equal. Moreover, data also has a lifespan. There might be data that is duplicated, outdated, and unnecessary. Don’t use ROT data for business insights. Keep only the data you need and protect it.



Forward-thinking organizations have an automated data management strategy in place that helps de-identify, manage, and enrich the data (making the data updated and accurate), grant access control, and archive and dispose of data.

Make your data management strategy easy to explain to regulators because, unfortunately, data breaches will happen. However, if you are properly safeguarding PII data and adhering to data privacy regulations, you can quickly respond to regulators' questions, such as "When did the breach happen? What was exposed? Who was impacted?"

According to the *IBM Cost of Data Breach Report 2023*, key categories of compromised records include customer PII with an average cost of $183 per record, and employee PII costing US $181. Within this model, a SharePoint site containing 22K customer records represents ~US$4 million in financial exposure.[3]

IBM, Cost of a Data Breach Report 2023.

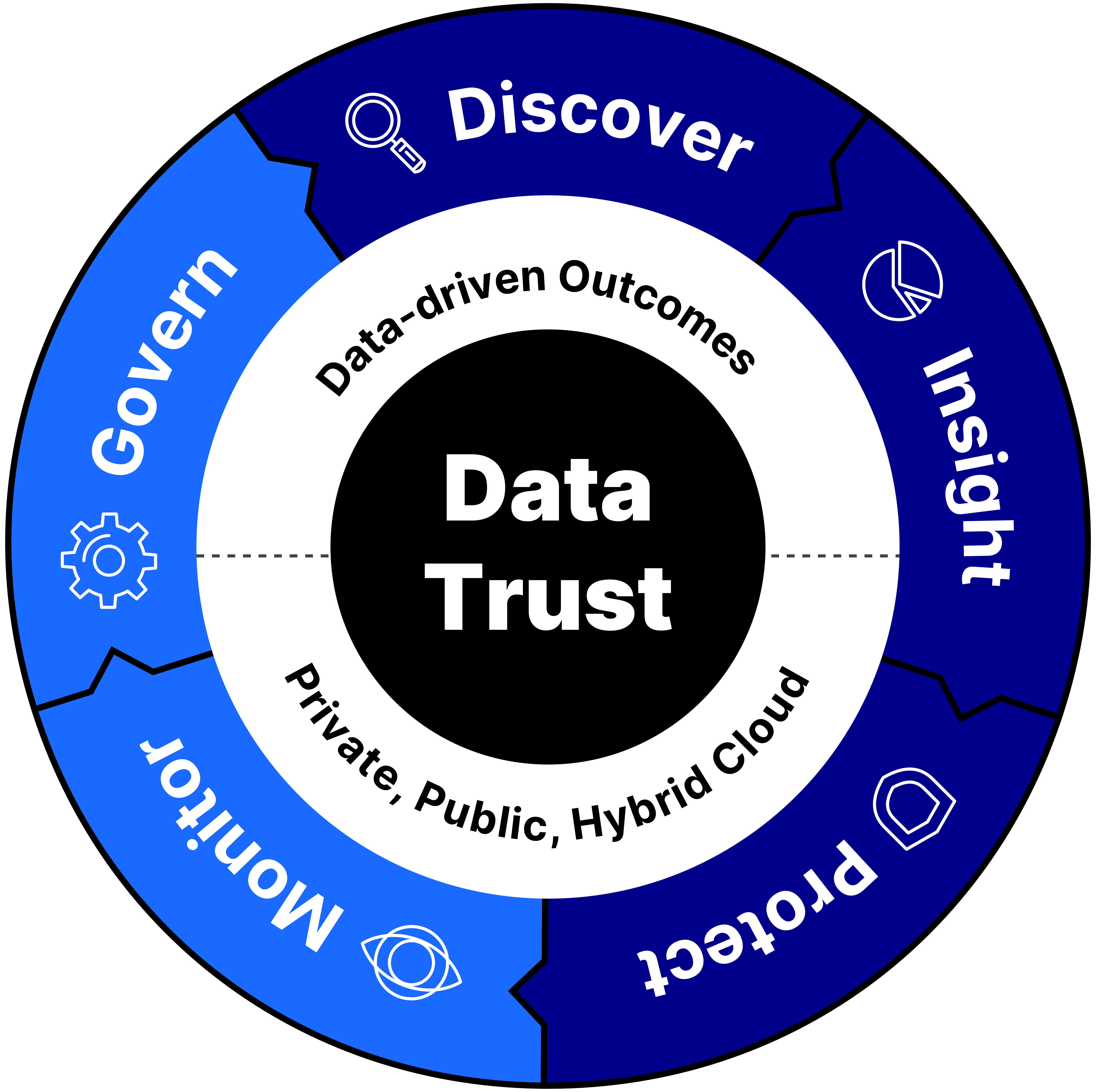
Strategy #4

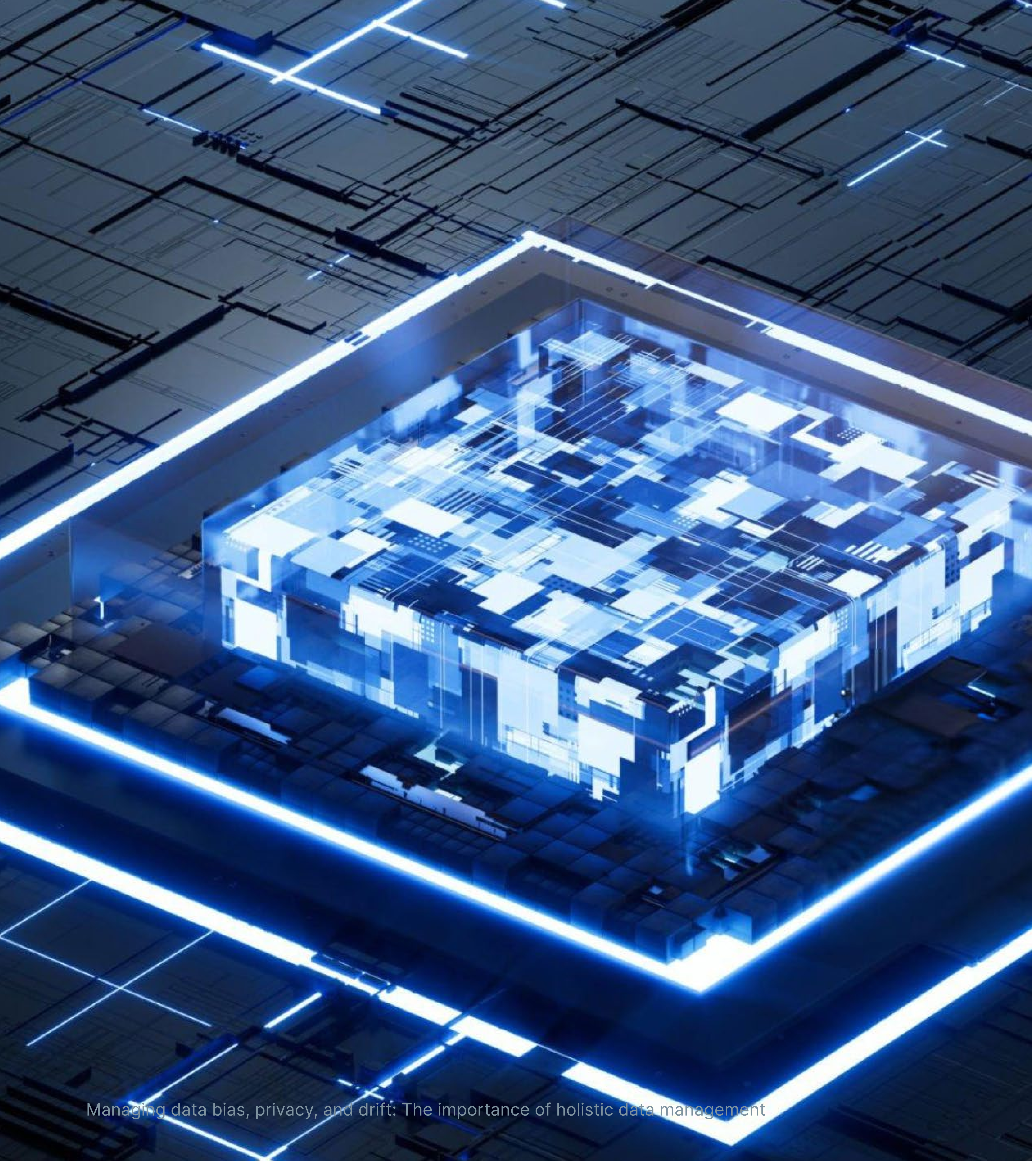
Effectively manage data access and compliance reporting to excel operational integrity.

Organizations must set the guardrails of who is accessing sensitive data and at what level. Safeguarding data access is vital, but it’s also critical to identify potential vulnerabilities—for internal security teams and external regulators.

Questions arise:

- Who is accessing the data?
- What level of access do they have?
- Is that data still valuable?
- Do you need to act on that data?





Consider a comprehensive data management strategy that pinpoints where valuable data is stored and identifies the individuals who can access them.

Relocate those files to an enterprise-grade content management system that augments levels of visibility and control.

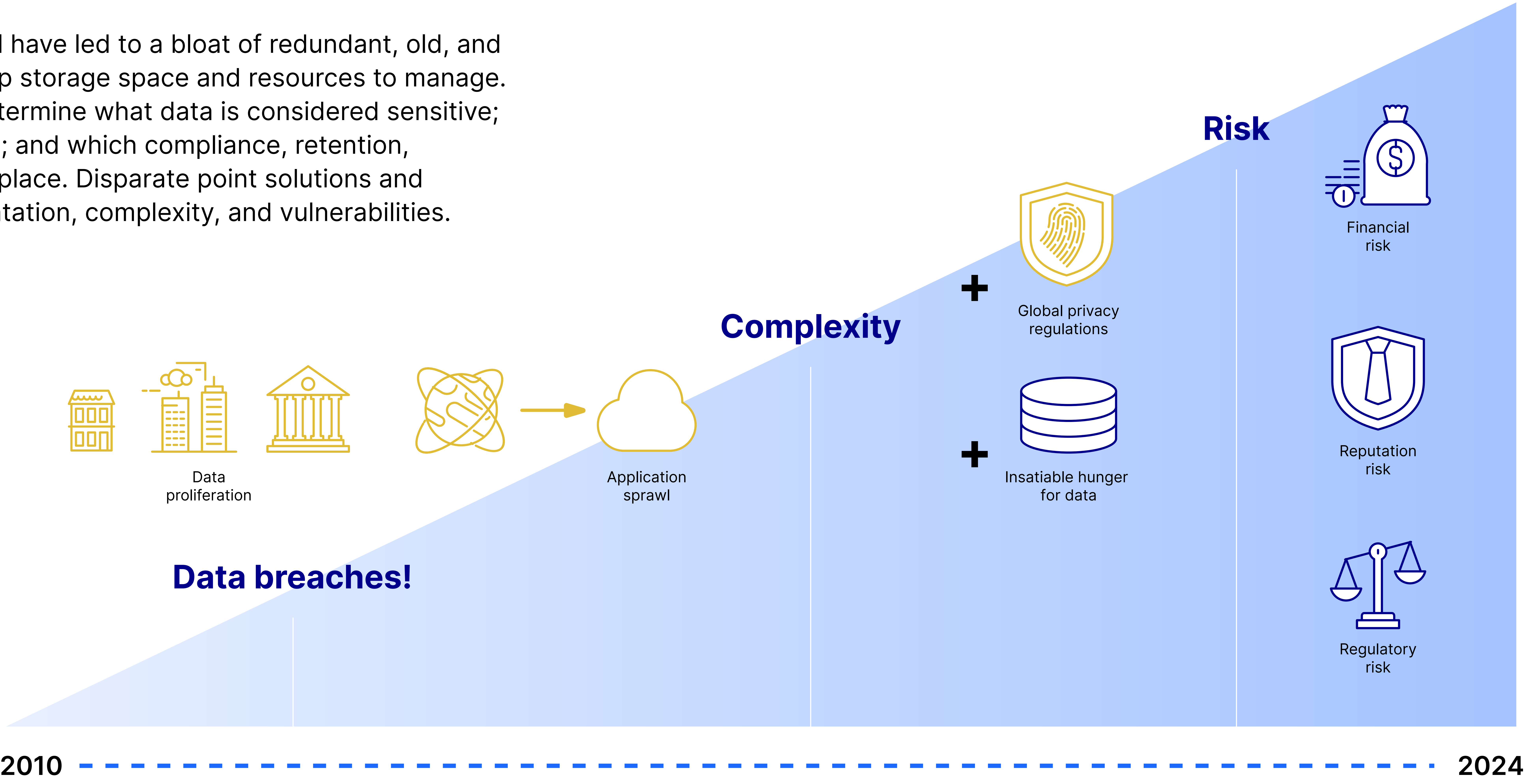
A managed data repository should also secure unstructured data and include data termination.

Overall, these granular approaches to data access management and compliance reporting can better position you to respond to regulators and auditors. You can provide concrete information on where you keep the data, how you categorized it, who can access it, how you secured the data, and when you will delete the data.

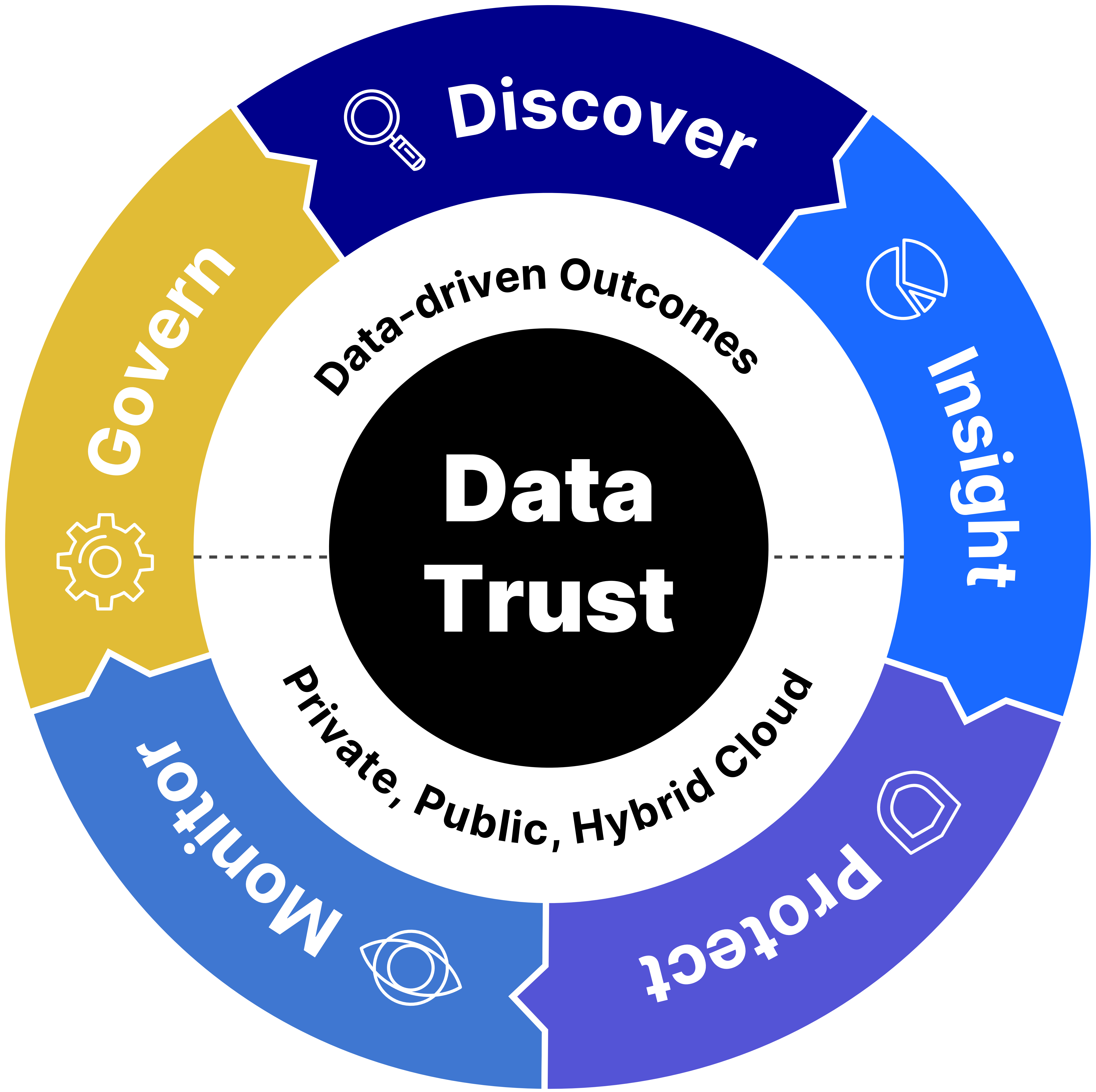
Strategy #5

Establish a robust data lifecycle management system for maintaining data relevance and minimizing risk.

Data and application sprawl have led to a bloat of redundant, old, and low-value data that takes up storage space and resources to manage. Organizations must also determine what data is considered sensitive; who has access to the data; and which compliance, retention, and security policies are in place. Disparate point solutions and approaches create fragmentation, complexity, and vulnerabilities.



Secure data management requires a holistic approach to overcome all the challenges covered in this report. A well-structured, holistic data management strategy includes data discovery and classification, gleaning accurate and actionable data insights, proactive data protection measures, data access monitoring and reporting, and solid data governance.



Conclusion

Combating data and application sprawl doesn't happen overnight. However, it doesn't have to be so challenging when you have the right strategy and technology partner. Organizations who embrace a holistic data management approach see the complete journey of the data while ensuring legal and record compliance, establishing the business value of the data, securing data, gaining customer trust, and augmenting data visibility and control.

OpenText has the experience, continuous technology innovation, and expertise of working with organizations to deploy a holistic approach to secure information management. With the right technology partner in place, organizations stay resilient, agile, and competitive today and in the future.

Learn more about our solution for data management.

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit opentext.com.

opentext.com | X (formerly Twitter) | LinkedIn | CEO Blog

opentext™