data.world

The Big Book of Business Glossaries

Insights on building a common language for your data-driven organization.

00.

Table of contents

01. Introduction 03

02. Challenges 04

03. The five pillars of effective business glossaries 09

04. How business glossaries work 11

05. How business glossaries help your business 12

06. Get started with business glossaries 13

07. The business glossary checklist 15

08. Start building semantic clarity 16

01. Introduction

The rise of the business glossary

More data means more potential for insights into your business. But more data also introduces a heightened risk of misunderstanding, inconsistency, and confusion across teams. When terminology isn't standardized, your business suffers from the consequences of what we call "semantic drift" — moments when your teams are using different definitions, making decisions based on different assumptions, or otherwise working at cross purposes.

In this ebook, we'll cover the challenges that semantic inconsistency presents, what business glossaries are and how they can help, best practices for improving data literacy and consistency, and how to implement business glossary tooling within your data tech stack.

The consequences of semantic drift can be severe, including wasted resources, inconsistent reporting, misaligned strategies, and loss of business trust. Research suggests that companies spend up to 30 percent of their time reconciling different interpretations of business terms, and 1 in 4 companies have made poor business decisions due to terminology confusion.

Fortunately, a powerful approach to data consistency and trust has emerged to address it: business glossaries integrated with knowledge graph technology.

02. Challenges

Expanding data ecosystems and constant evolution

Modern data environments are growing in complexity and constantly evolving. With the widespread adoption of cloud services like Snowflake and Databricks, the data ecosystem now involves more sources, more stakeholders, and more opportunities for misinterpretation than ever before.

The old data governance principle of "define once, use everywhere" still holds true, but is no longer sufficient. Semantic issues can occur at any stage of the data lifecycle — from ingestion to transformation to analysis to presentation. Teams can be blindsided by inconsistent definitions or confused context that goes undetected for months — until a critical meeting where executives are looking at the same metrics but seeing different numbers.

The constant evolution of business terminology adds more complexity. As your organization grows, mergers happen, or new products launch, keeping business glossaries and data dictionaries up-to-date becomes increasingly difficult.

Without clear visibility into how terms are defined, used, and related to data assets, resolving semantic disagreements in a complex, dynamic environment becomes increasingly burdensome.

CASE STUDY

OneWeb faced the challenge of managing massive volumes of satellite data—55 billion rows daily across 32 Snowflake tenants—making data discovery overwhelming for employees who needed information outside their immediate domain. Using data.world, OneWeb implemented a comprehensive data catalog with semantic search capabilities and lineage tracking that allowed users to find and understand data without direct access to all systems. Miguel Morgado, Product Owner at OneWeb, noted that "data.world is a fantastic tool for the data mesh approach," which helped the company drastically reduce data sharing timelines from weeks to seconds, accelerate dashboard creation, and successfully implement a full data mesh with just seven team members in under a year.

OneWeb gets data from space to Earth's end users in minutes.

02. Challenges

Growing teams and lack of standardization

Along with data ecosystems, modern data teams are growing in size and complexity. This means more people are working with your data, bringing their own terminology and assumptions.

For example, today's data teams often include data engineers, data scientists, data analysts, and analytics engineers. As those teams mature, decentralized organizational models will give ownership of analytics to specific departments. This can lead to confusion about what terms mean, inconsistent definitions, and a lack of perspective on how semantic drift impacts the entire business. The potential for complication only increases for advanced teams adopting data mesh frameworks, which distribute ownership of data across domains.

Decentralized teams can be incredibly effective if terminology is standardized and definitions are clear. But when team members lack a shared understanding of what business terms mean, seemingly minor definitional differences can cause problems like contradicting insights, conflicting courses of action, and essentially - more semantic drift.

CASE STUDY

Learning Care Group, a leading provider of early childhood education, was dealing with inconsistent definitions across their hundreds of locations. "We're data rich. But how do we get people to use our data more? How do we get them to access the data, understand it, and then build off it?" said their manager of BI and Analytics Delivery. Data.world's consumer-grade UX allowed Learning Care Group employees to search across discussions, analysis, and more to easily find and understand the data they need.

Learning Care Group uses data.world to provide data access, business context, and inform employee decisions.

02. Challenges

No central source of truth

Traditionally, data teams have used spreadsheets or wikis to document business terms and definitions. But today, companies manage so much data that these manual approaches can't scale.

Spreadsheet-based glossaries or static wiki pages can document specific terms with definitions, but achieving organization-wide adoption and maintenance is a significant challenge. These approaches often become outdated as soon as they're published, and keeping them current requires dedicated resources that often get deprioritized against more urgent requests.

And even well-maintained static glossaries can't solve the deeper problem of connecting terminology to actual data assets. There's a critical gap between defined terms and their implementations in reports, dashboards, and data models. The glossary might say one thing, but the actual implementation does something else — and nobody notices until critical decisions have already been made based on faulty assumptions.

Penguin Random House

CASE STUDY

Publishing giant Penguin Random House needed to provide their business users a single reference for the location of any data set, and a single source of truth for what that data means."Imagine managing books without title information, author data, cover images, royalties, or number of chapters. That's what it's like managing data without a catalog. Now, information that once took our data scientists a week to find is discoverable in seconds on data.world.

How Penguin Random House takes a card catalog-like approach to data

02. Challenges

AI readiness demands semantic clarity

With the rise of generative AI and large language models, the need for well-defined business terminology has never been more critical. AI systems are only as good as the data and context they're given, and semantic inconsistency can amplify issues dramatically.

When organizations deploy AI solutions without addressing underlying semantic problems, they risk automating and amplifying misunderstandings. AI systems trained on inconsistent definitions will produce inconsistent outputs, potentially at scale and with greater confidence—making the problem worse, not better.

And if your teams can't agree on what key terms mean, how can they possibly evaluate whether an AI system is producing correct and valuable results? Without a foundation of shared understanding, AI initiatives are built on shaky ground.

For organizations seeking to become AI-ready, establishing business glossaries and semantic standards isn't just good practice, it's an essential prerequisite for success.

NCL NORWEGIAN CRUISE LINE®

CASE STUDY

For Norwegian Cruise Lines, legacy systems from different decades created a complex web of data that made it difficult to leverage insights for customizing the guest journey. With over 90% of its 40,000 employees working onboard ships where connectivity is often limited, NCLH needed a solution to centralize its guest data and streamline operations across its 28 ship fleet.

Enter the data catalog: "Most importantly, we can demonstrate what the data is doing and how the LLM is generating responses—this level of explainability significantly improves trust and accelerates the adoption of AI," said Cris Hadjez, Senior Director of Data Governance

How Norwegian Cruise Line Holdings reduced issue resolution time from hours to minutes.

02. Challenges

Losing organizational trust in data

Every year, companies are investing more money in data and touting their status as data-driven organizations. 97.2% of companies are now investing in big data solutions, and 87.9% of organizations prioritize data and analytics investment.

But when semantic drift occurs, trust in data erodes. Business leaders spend more time debating what terms mean versus acting on insights, and the old habits of relying on gut instinct or intuition start to creep back in.

This makes sense: if business leaders can't trust that everyone is talking about the same thing when they use a term like "active customer" or "qualified lead," their skepticism is warranted. And even if data teams address these problems and implement better glossaries and processes to establish semantic consistency, the damage is already done.

PROLOGIS

CASE STUDY

For Prologis, the world's largest logistics real estate company operating across four continents and 19 countries, maintaining trust in data became critical. Historical views were no longer sufficient for future growth. They implemented data.world's knowledge graph platform to combat semantic drift by establishing clear relationships between data elements, business processes, and decision-making frameworks.

The centralized platform enabled anyone in the organization to access well-organized, clearly explained data, creating a foundation of shared understanding that proved particularly valuable as the company expanded its AI/ML initiatives. "Where data really becomes powerful is when you can tie data to business processes, and business processes to decision-making processes within the organization," explained Slotwinski. Their semantic clarity approach has become the cornerstone of their advanced analytics strategy.

Prologic catalogs people, processes, and decisions.

03. The five pillars of effective business glossaries

A business glossary is a centralized repository of business terms and their definitions, providing a common language across an organization. Its goal is to reduce the frequency and impact of semantic drift, and it's an important part of overall business literacy.

Drawing on best practices of data governance, modern business glossaries go beyond simple term definition to establish rich context and relationships between terms, data assets, and business processes.

The five pillars of effective business glossaries are:

- **Clarity:** Are definitions clear, concise, and free of jargon?
- **Consistency:** Are terms used consistently across systems and teams?
- **Context:** Do definitions include business context and examples?
- **Connection:** Are terms linked to related data assets and concepts?
- **Curation:** Is there clear ownership and a maintenance process?

> "You hold all of this institutional knowledge, you sort of know how things work, how we're encoding business logic and the value of the data that you have access to, but all of your friends scattered around the business really don't. It's a big job translating, making that accessible and trying to bridge the gap."

Leah Weiss
CEO/Co-founder
PREQL

04. How business glossaries work

Business glossaries reduce semantic drift by providing a central repository of standardized definitions connected to the data assets where those terms are used.

Modern, knowledge-graph-powered glossaries go beyond simple definition lists to create a rich semantic layer over your data ecosystem.

Centralized definitions reduce misinterpretation

Modern business glossary tooling provides a single source of truth for business terminology. Your team establishes clear, consensus-driven definitions that reduce misinterpretation and standardize communication.

Knowledge graph connections enable context and discovery

With the semantic relationships provided by knowledge graph technology, teams can visualize how business terms relate to each other, to data assets, and to business processes. Your team understands the full context of a term, and can discover related concepts and data sources.

> "ETL only exists because of a lack of interoperability. Seamless exchange of data is really hard because formats don't work together as they should. That's one of the big pain points. The problem statement is how do you make the seamless flow of data from one bounded context to another? If we solved for that, we would all be working on much more interesting higher order problems."

Dael Williamson
EMEA CTO
databricks

05. How business glossaries help your business

- **Decision-making speed and quality**
  Clear, consistent terminology carries real value. Reducing semantic drift means tangible benefits for decision-making. Think about the time wasted in semantic debates; they don't happen when everyone agrees on words, acronyms, and context.

- **AI readiness**
  The knowledge graphs that power modern business glossaries provide the semantic foundation that AI systems need to function effectively. By establishing clear definitions and relationships between terms, glossaries give AI systems the accurate context they need to work. This semantic foundation is particularly critical for AI-ready data. Organizations with well-maintained business glossaries find their AI initiatives deliver value faster and with fewer false starts.

- **Data literacy**
  Business glossaries serve as powerful educational tools, helping new team members quickly understand your organization's terminology and how it relates to your data and business processes. This accelerates onboarding and improves data literacy across the organization. Glossaries make data more accessible to non-technical users.

- **Semantic health**
  Modern business glossary tools make it simple for teams to measure semantic health across the organization. These metrics typically include coverage (percentage of critical terms defined), adoption (how widely the glossary is used), and connection (how comprehensively terms are linked to data assets). Tracking these numbers helps manage stakeholder expectations. Data governance leaders can quantify and demonstrate the value their glossaries are providing to the organization.

> "What does it mean to be data-driven? It means relying on data to make decisions rather than on a gut feeling. But if you don't know what data in the enterprise is available, how can you make data-driven decisions? And how can you be AI-ready if you don't really know what you use to train your AI?"

Olga Maydanchik
Data Management Practitioner
Anywhere

PODCAST
Want to Be Data and AI-Driven? Start with the Basics…a Glossary! with Olga Maydanchik

06. Get started with business glossaries

When you introduce any new element into your data governance program, it pays to adhere to best practices—and business glossaries are no exception. As you adopt glossary tooling, mind the following principles:

Start small and focus on impact

Don't try to define every term in your organization at once. Begin with the most critical business domains or the areas experiencing the most confusion. Define the terms that matter most—those used in key reports and decision-making processes—and expand from there. Success with high-visibility terms will build momentum for broader adoption.

Establish ownership

Before building a business glossary, it's critical to establish who owns the definitions and how they'll be maintained. Each term should have a designated owner who is responsible for its definition, and there should be clear processes for requesting new terms, updating definitions, and resolving conflicts. Without this governance infrastructure, glossaries quickly become outdated and unused.

> "I like to say everybody is data literate. We use data all the time. But what's happened is, we live in a very distracted world, in that there are distractions everywhere. We take in information left and right. Instead of questioning it, we take it at face value and move on. I see organizations that are doing really well with it, but it definitely has a long way to go."

Jordan Morrow
SVP of Data and AI Transformation
AgileOne

06. Get started with business glossaries

Connect glossary to data

A standalone glossary has limited value. The real power comes from connecting terms to the data assets where they're implemented and the business processes where they're used. This connection transforms a simple list of definitions into a navigable knowledge network that provides context and enables discovery.

Build a collaborative culture

Business glossaries aren't just IT or data governance tools. They're organizational assets. Involve subject matter experts from across the business in defining terms, and make the glossary easily accessible to everyone. The more people who contribute to and use the glossary, the more valuable it becomes.

Treat it like a product

Business glossaries, like any other data asset, should be treated as a product, not a one-time project. A few tips include aligning your glossary goals with the goals of the business, incorporating glossary development into your data governance roadmap, and setting usage metrics to measure the effectiveness of your glossary tooling.

07. The business glossary checklist

To achieve the goal of reducing semantic drift, your business glossary should include, at a minimum:

- Clear, concise definitions written in business language
- Designated owners for each term
- Version history and audit trails
- Examples of proper and improper usage
- Related terms and synonyms
- Connections to data assets where terms are implemented
- Business context including where and how terms are used
- Classification of terms (e.g., by business domain)
- Search and discovery capabilities
- Integration with data catalog and metadata management
- Collaborative features for feedback and discussion
- Governance workflows for approval and updates

08. Start building semantic clarity

Business glossaries help teams build trust in data by eliminating semantic drift and increasing terminology consistency. And with the data.world platform, your team could start reaping the benefits of knowledge graph-powered glossaries immediately.

- The Data Prose Blog
- The Big Book of AI ready data
- The Catalog & Cocktails podcast

data.world