OCTOPAI

The Automated Business Glossary:

Everything You Need to Know

eBook

If you're building a business glossary that isn't automated, you're wasting your time. You read that correctly. **YOU'RE WASTING YOUR TIME.** Why? Because getting a business glossary up and running can take your team months, but an Automated Business Glossary takes only a few short minutes.

And time is money, people.

Business Glossary: The Need

So many errors, false assumptions, and general misunderstandings result from ambiguous terminology. When team members incorrectly assume that everyone has a shared understanding of the terms being used, chaos can ensue. This is why most contracts have a section right at the top that defines the terms used in the document, even seemingly trivial terms such as "us" and "you."

A common understanding of the terms used in a business is essential, especially in large, diverse, global companies where not everyone has the same cultural or business background. The key to this common understanding is what's known as a *business glossary*.

What Is a Business Glossary?

We all know what a glossary is: An alphabetical listing of terms and their definitions, usually consisting of the terms used in a book or by a particular field of interest. A business glossary is similar—in essence, it's an alphabetical listing of terms and their meanings as used in a specific business. It provides an authoritative source of meaning for the terms used in all of a company's operations.

It sounds simple, right? After all, everyone knows what a "customer" is, so why do we need a formal definition for it?

Dig a little deeper in any good-sized company, however, and you'll see why it's needed. Depending on where the data is coming from, fields that mean the same thing could have completely different names, or the same concept might be represented in different ways. Even something as simple as a person's name can get complicated; in different systems, it might be represented as:

- A single "Name" field, formatted *first_name last_name*
- A single name field, formatted *last_name, first_name*
- Separate fields for first and last names, with a wide variety of names and labels: firstName, Name_F, FName…

OCTOPAI

Throw in data sources from foreign countries, where field names and labels might be in foreign languages, and the possibilities are quite literally endless. Having a business glossary ensures that a term means the same thing across all departments, business units, and geographies. It serves to reduce or eliminate confusion or assumptions regarding the meaning of a term as applied in the business. What does any of this have to do with business intelligence, databases, data warehouses, and the like? Plenty.

Consider two reports, one used by the marketing department and the other by the technical support team. Each report has a column for the customer's name, but each column might be labeled and formatted differently. Or, one might have separate columns for first and last name, whereas the other one has a single column with first and last name concatenated together. What enables these separate concepts of "name" to be matched up when data from each application is loaded in a data warehouse? Metadata.

When fields in different systems have different names for the same thing, the only thing that sets them apart from each other is the metadata associated with them. The business glossary serves as the source of truth for the metadata.

Thus, a business glossary isn't just a convenient listing of terms to keep everyone on the same page. It's an important reference for business analysts, data analysts, data scientists, data governance managers, and everyone on the BI team.

How do they use the business glossary?

- **Consistency** - A good business glossary helps data professionals ensure that data assets, reports, and dashboards represent data consistently, both within and among the various objects.
- **Management** - A common understanding of business terms makes it easier to create, maintain, and integrate new sources of data in the environment.
- **Development and delivery of new reports and dashboards** - Standardization enables data professionals to make the connections needed between data elements that mean the same thing but have different names.

In short, a business glossary is useful not only to the day-to-day business users but also to the BI technical staff and related roles.

Side note: Don't confuse the term "business glossary" with "data dictionary." Although both are based on metadata, a data dictionary is a low-level reference describing the attributes of columns and tables in a database. It has little to do with the higher-level business terminology. Business users will consult the business glossary but will probably never see a data dictionary.

Building a Business Glossary

To build a business glossary, a good place to start is existing reports, databases, and data warehouses. Why? Because these are the places where important business terms end up. If "customer," "cost of goods sold," or "full-time employee equivalent" are important terms for a business, they're going to be represented in reports somewhere.

The trouble with this approach is that it is a manual, tedious, painstaking process that, depending on the size and complexity of the business, can take a whole BI team or professional services group anywhere from several weeks to several months. Simply tracking down and interpreting all the metadata requires a significant effort by a small army of analysts. It's a process fraught with frustrations, such as:

- Incomplete, incorrect, or obsolete metadata
- Data transformations and calculations that don't mean what people think they do
- Conflicting assumptions within the business regarding the meaning of a given term
- Time-consuming for many different people in the organization

There could also be a fair amount of detective work involved, or something akin to archaeology. Some of the data sources could be from 10 or 20 years ago, if not longer, and the people who originally worked on them are now long gone, leaving little or nothing in the way of documentation.

Too many data sources are notorious for their lack of rigor around consistent data typing, calculations, and metadata. In some cases, figuring out what certain things mean might be left to making a best guess.

Happily, there's a better way: **Automation.**

Taking the Business Glossary to the Next Level With Automation

Building a business glossary the old-fashioned way can become mired in the details, intricacies, false leads, and dead ends of your labyrinthine data environment. In doing this manually, many BI teams get lost in the weeds and eventually get burned out on pursuing a project that doesn't seem to make much progress over time.

To shorten the project cycle and take your business glossary to the next level, you can have your business glossary generated automatically.
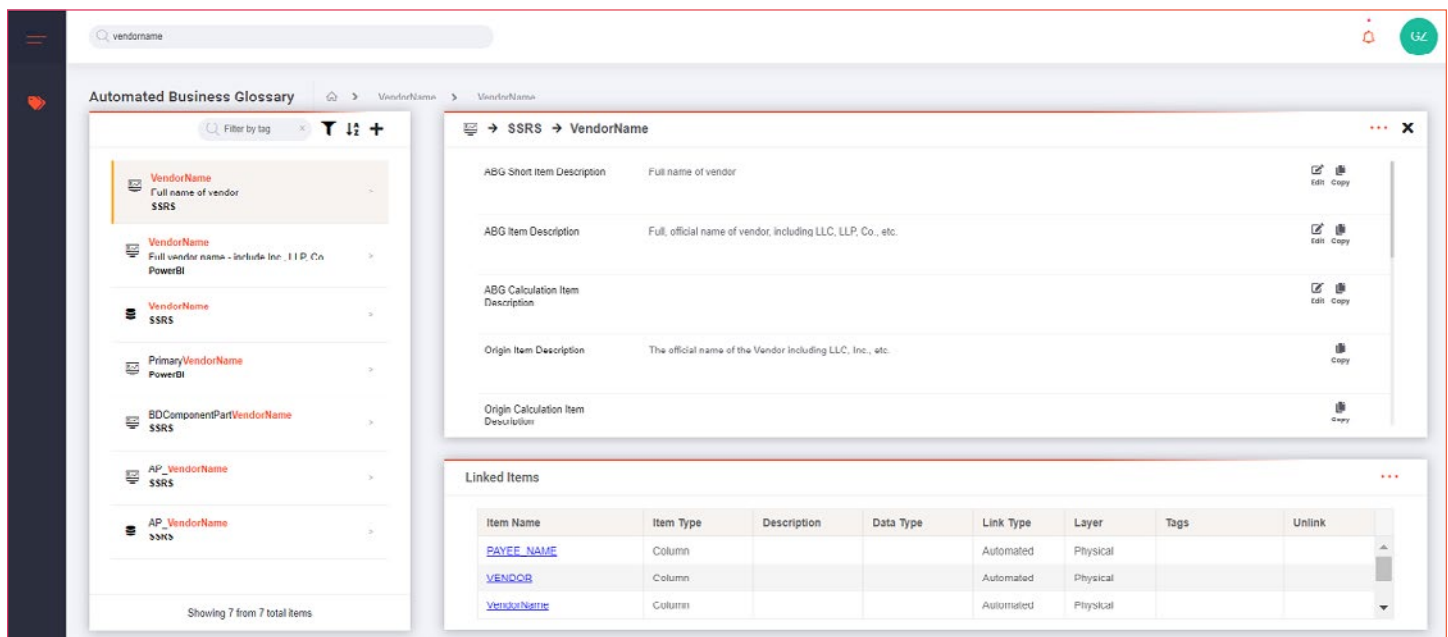

Figure 1: Results generated within Octopai's Automated Business Glossary after searching the term "vendorname."
Left panel: Search results of all places this term appears within physical, semantic, and presentation layers of each reporting system.
Top right panel: All the attributes for the selected term (in this case VendorName in the presentation layer in SSRS) such as Descriptions, Source System, Owners, Tags and more.
Bottom right panel: Linked items for the selected term which help users understand that those items have the same meaning even if they don't have the same name.

Generating your business glossary automatically will simplify and streamline the execution of the following tasks:

- **Centralize data layers from all reporting systems:** In this context, the term data layer refers to any component of the data structure model that includes the physical, semantic, and presentation layers:

  - **Physical layer:** This layer includes the actual tables and columns that define the data structure at the database level.

  - **Semantic layer:** The semantic layer represents the data from the user's perspective. It hides the details and complex relationships of the underlying physical layer and packages them into data objects that correspond with the organization's understanding of the information in common business terms. (This is one of the reasons why a solid business glossary is so important.)

  - **Presentation layer:** This layer is what shows up on columns in reports and dashboards, and it often involves some kind of summarization or aggregation into a small number of key performance metrics. It can also be in multiple languages.

- **Standardize data layers across different reporting systems:** Centralizing the data layers involves standardizing them across all of these systems.

- **Include logical and calculated data items:** Logical data items are not represented in a single column, table, or data source; a logical data item is composed of data items transformed and combined from other sources. Their close cousins, calculated data items, are calculated from multiple other numerical data items using a defined formula. Both of these must be accounted for when building a business glossary.

- **Reduce or eliminate large-scale data entry projects:** No one wants more tedious grunt work, and there are few activities as mind-numbing (or error-prone) as manually entering data, especially data that already exists somewhere in some form.

The key to success for all of these tasks is automation.

What Happens When You Automate a Business Glossary?

A business glossary can offer the following benefits:

- **One source of truth:** Having a single source of truth for your company's metadata enables you to build better, more consistent reports and dashboards and gain more control over your data assets. In turn, this enables better enforcement of data governance policies and procedures.

- **Common business language:** Having a common business language means that people in different departments and business units have a common understanding of every important term, and nothing gets "lost in translation." This can prevent embarrassing internal or public miscommunications.

OCTOPAI

- **Improve self-service BI:** The modern trend of enabling self-service BI provides great benefits by reducing the load on the BI staff and improving the turnaround time for desired reports and dashboards, but it can be a disaster without a common business language. A business glossary means that a piece of information (a report column or key performance metric) means the same thing no matter what report it appears in or who created it.

But when you apply automation to generate a business glossary from your company's metadata resources, the benefits grow profoundly:

- **Time and money savings:** The ability to build a business glossary automatically has the obvious benefit of reducing the time and labor costs required to complete the project. It also has the extra added benefit of reducing wear and tear (read: burnout) on your BI team and the rest of the organization that is involved.
- **A snapshot in time:** As noted earlier, constructing a business glossary manually can take a long time, and to keep everything straight during the project, you have to either freeze all the data assets (i.e., prevent adding new assets or deleting or modifying existing ones) or try to keep up with changes as they happen—neither of which is especially practical. With automation, you can get a "snapshot in time" to serve as your baseline and then continually monitor for changes and update the baseline.
- **Increase report accuracy and reduce errors:** Automating the business glossary reduces the chances for error inherent in manual processes, meaning that the reports that are built on the basis of the terms in the glossary will be more accurate and have few, if any, errors.

The overarching benefit to a business glossary is increased company-wide consistency and data integrity, and using automation will turn this enormous project into a task on a to-do list. When new businesses are acquired, it becomes easier to integrate their data assets and reports, because you have an established standard that makes it easier to "connect the dots" with the new data sources. Keeping up with changes to existing data assets is easier as well because the automated process can be run at any time—and should be run periodically, so there are no surprises.

The Time to Automate is Now

The time to decide to automate your business glossary is now—before you embark on a business glossary project. Automating your business glossary pays for itself many times over—not just in the initial business glossary project but in the ongoing maintenance tasks to keep the glossary up to date.

Octopai's metadata management automation platform can be the cornerstone of your business glossary project. With Octopai, you can be set up and starting the process of building your business glossary within one business day, eliminating the manual drudgery and freeing your BI team up to work on deriving insights—the work they were hired to do.

In addition to importing all metadata items from the reporting systems, we also import the associated original descriptions for those items, instead of doing it manually. Huge savings.

If you are planning, starting, or restarting a business glossary project, automation can help get your project across the finish line much faster and more accurately, and bring your company the consistent business terminology it needs to succeed in no time.

OCTOPAI