

EBOOK

An Expert's Guide to DATA ENRICHMENT



For years, organizations have invested in gathering and using the data they generate from internal operations. And to be fair, they've made considerable strides to realize greater value from their internal data.

But today, data teams are encountering another challenge: their internal data is incomplete, inconsistent, and unclear. And it's no surprise that this is the case. Internally, data lives across a myriad of disparate systems across multiple departments. And the data within these systems mirrors the way that part of the organization works, which often doesn't align with the practices and processes in other departments.

This challenge comes to a head as organizations look to democratize their data and analytics. They spend a great deal of energy and effort implementing self-service BI tools such as Qlik, Tableau, or PowerBI only to realize that while they've fixed the veneer, they've exposed their [data's dirty, rotten core](#).

Democratizing data: providing everyone, regardless of

their role, a self-service way to easily work with the data, understand the data, and confidently use it to make data-driven decisions.

But here's a secret that many organizations have only recently discovered: the best version of their data doesn't reside within their company walls. Instead, it lives externally, within an outside source such as ZoomInfo, Clearbit or [Dun & Bradstreet](#).

Think about it: does the best, cleanest version of your personal data - your address or phone number, for example - live within the ERP or CRM systems of your service providers? Most definitely not.

To overcome challenges with unclear and incomplete data, organizations need to look beyond their own internal data sources. They need to link their valuable, internal data with external data provided by third parties, vendors, or public data sources through data enrichment so that they can provide decision makers with data that is cleaner, more accurate, and more complete.

What is data enrichment?

Data enrichment is the process of enhancing existing, internal datasets with information that is generated from additional data sources. These sources could include data about organizations, people, or parts or data that could be used for sales and marketing, product analytics, risk management, and more.

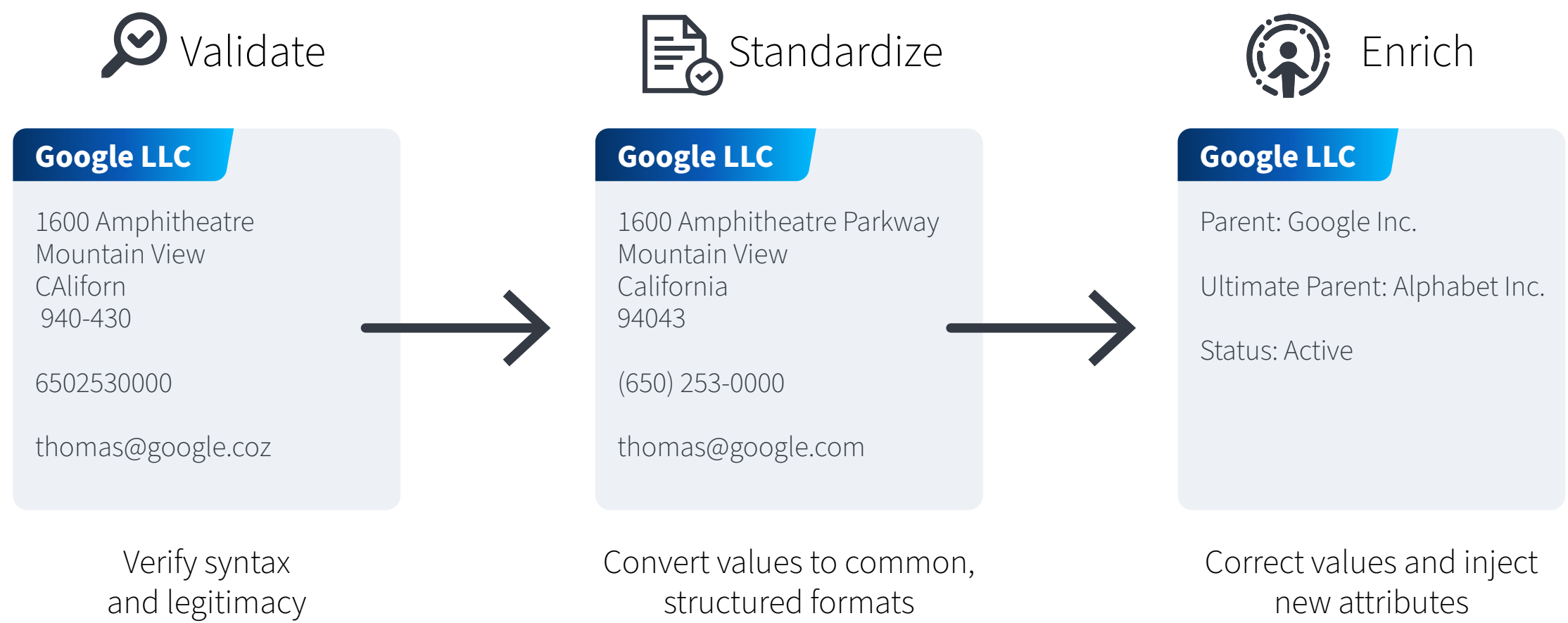
Three different types of data enrichment exist: validate, standardize, and enrich. Each one builds on each other to unlock greater value from your organization’s internal dataset. Here is what each

type looks like using the corporate information of Google LLC as an example.

- **Validate** verifies data syntax and content legitimacy. For example, this step helps verify whether the email domain of Google LLC is legitimate, whether or not the address actually exists, and whether or not the phone number is a landline or mobile.
- **Standardize** converts values to a common and structured format. For example, this step

standardizes the phone number and the addresses into a standard format. This step also corrects spelling mistakes if they exist.

- **Enrich** corrects incorrect values and injects new attributes. For example, this step tells us that Google LLC has an ultimate parent company that is Alphabet Inc. This step can also provide other attributes for analytics purposes, such as revenue or number of employees. The potential use cases for analytics can be infinite.



With all three types of enrichment, you can gain a full, holistic view of the entity in order to make the best business decisions.

Why businesses use data enrichment

Data enrichment helps companies create and realize business value in numerous areas, including use cases such as:

- **Customer analytics:** with data enrichment, you can better understand customer behaviors and target the right prospects. For example, by using firmographic data, employment growth, and technologies that companies use, you can identify ideal B2B prospects and their lookalikes for your products.
- **Strategic decision making:** data enrichment helps you to identify major organizational improvement opportunities by benchmarking external data. For example, you can conduct peer analysis on talent profiles and job postings to determine how to better retain top talent.
- **Operations forecasting:** with data enrichment, you can forecast key metrics and economic factors in your business for operational decision making. As an example, based on demographic shifts and other private company data, you can forecast which customer segment is more likely to grow.
- **Risk management:** data enrichment enables you to identify key operational, financial, and reputational risks by identifying who you are dealing with. For example, you can avoid supply chain disruption by identifying the parent companies and seeing if they are sanctioned or in certain geographical locations that expose you to certain risks.



Overcoming obstacles to data enrichment

Enriching your internal data with data from external providers is a surefire way to improve its quality. However, many organizations encounter three primary obstacles as they look to realize value from external data.

- 1. Knowledge of data availability:** it's difficult for data teams to gain a basic understanding of what data is available. Why? Because the ever-expanding volume and variety of external data is expanding quickly, making it difficult to keep up with the pace of change in the data sources.
- 2. Fragmented data sources:** data teams often struggle with the thousands of data products they can obtain through a multitude of channels. And the sources often come with heterogeneous formats and schemas to clean up as well.
- 3. Operationalize external data:** in order to efficiently use external data in operations and analytics, organizations may need to conduct a major upgrade to their existing data environment or workflow.

To overcome these obstacles, we believe there are five key questions you must ask - and answer - in order to realize the most value possible from external data.

1. What data attributes do you need - and how do you identify them?
2. How do you assess the sources you need?
3. How do you integrate the third-party data with your own, internal data?
4. How do you get value - and how do you measure it?
5. How do you make the process continuous to ensure you're always gaining value?

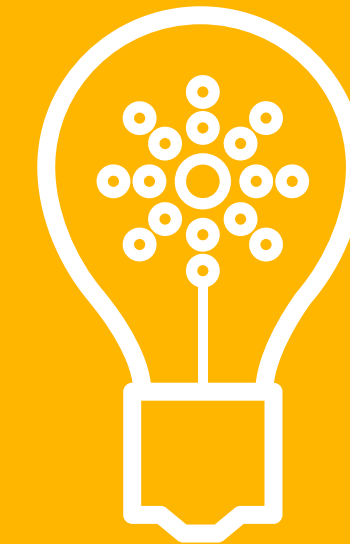
Let's take a look at each one in detail.



1. What attributes do you need - and how do you identify them?

When it comes to enriching data, the best place to start is with the business question. It's important that you understand and assess which questions your business leaders ask the most - and where they most-often get the wrong answer.

These questions form your hypothesis, helping you to determine which questions you need the data to answer. That way, you can ensure that the data you source will provide the insights your business leaders need to make better decisions.



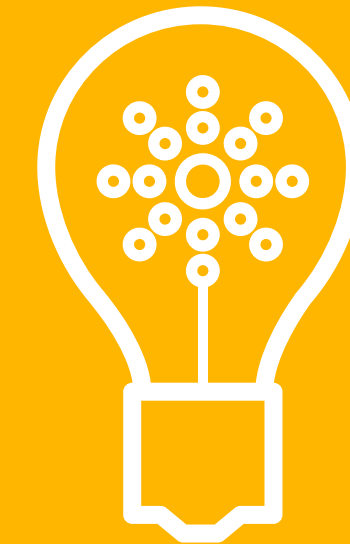
Pro tip:

Too often, companies make the mistake of talking with popular third party providers and assuming that once they have the data, the insights will come. This is simply not true. Instead, go into these discussions with your hypothesis on which questions you need to answer, and ensure that the third-party provider has the data you need to answer them. It's never a good idea to select a provider and assume you can define your use cases later.

2. How do you assess the sources you need?

Many third-party data providers talk about their quality and coverage rates. And while important, what's even more critical is the ability to link sources and attributes together.

Make sure you are asking your external data providers what level of detail they can provide. And then make sure they are providing what you need. It's a smart move to conduct a proof of concept or match test to ensure the vendor you're working with is a strong provider of the data you need.



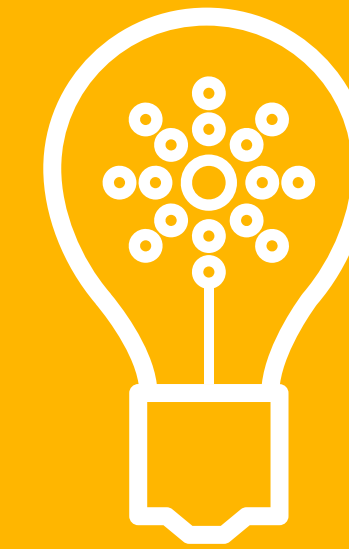
Pro tip:

When evaluating external data providers, don't fall into the trap of only relying on the big players. After all, just because they provide coverage on millions of companies doesn't mean they will provide the data your organization needs. In fact, many times, niche players provide more value, especially for data sets that are smaller, and more unique to your business.

3. How do you integrate the third-party data with your own, internal data?

Integration of internal and external data is a critical step. And to integrate your data successfully, you need an automated process for matching data and workflows that allow you to easily clean up unmatched records.

You also need stewardship over the data so that you get the enriched records to the point that you can master them. It's a process that requires both humans and machines, working together.



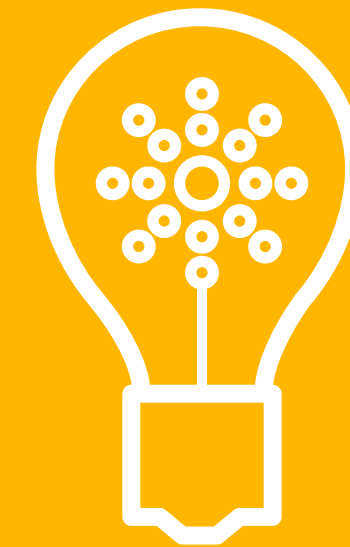
Pro tip:

Avoid relying too heavily on your third-party provider's match services. Why? Because if you use multiple data providers (and you should), each one with its own match service, you may end up with different results from each. Instead, bring the process in-house, and ensure your team becomes expert at the discipline of enriching data. In the long-run, having this ability, along with the right tools and technologies, is more cost effective for the business.

4. How do you get value from enriched data - and how do you measure it?

When you make an investment in data enrichment, it's important to show the value to your organization. By correctly defining the attributes you need, as we suggested in question number one, then you should be well on your way to gaining value from your investment. Measure how well the external data provider delivers on these attributes and the improvement in your data sources as a result.

Internal feedback is another way to assess value. Ask your business users for feedback - and ask them to be honest. That way, you can adjust your strategy if an external source isn't delivering its promised value.



Pro tip:

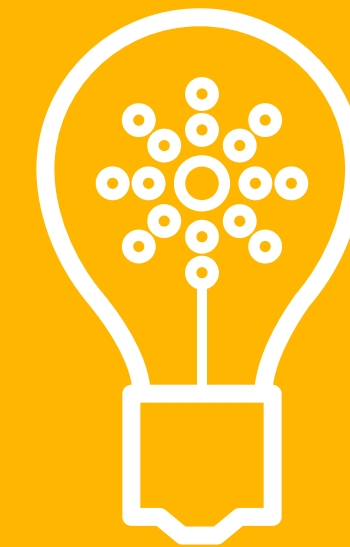
An often overlooked measure is time savings. Many people in your organization are likely using Google to enrich their data by looking up addresses or phone numbers on a company's website one by one. But this method of enrichment (obviously) doesn't scale. Using third party providers will enable you to enrich the data at scale, saving business users time and enabling you to gain an aggregate understanding of the enrichment process.

5. How do you make the process continuous to ensure you're always gaining value from your data enrichment efforts?

To ensure that you are continuously gaining value from your data enrichment efforts, there are two steps you must take.

First, set up a pipeline for continuous enrichment. Make it an ongoing process instead of a point-in-time effort. That way, you are continuously improving the quality of your data and, in turn, your analytics as well.

Second, be mindful of when your internal data begins to drift. Over time, new questions will arise that your business users must answer. And these questions could require enrichment from new external data sources.



Pro tip:

Because data is constantly evolving and changing, so, too, must your data enrichment strategies and providers. Keep your finger on the pulse of the business and you'll always stay one step ahead of your business users' needs.

Simplifying and Improving the Data Enrichment Process

As companies continue to spend significant amounts of time and money to derive business value from their disparate data sources, they need a data enrichment service that makes their data cleaner and more complete. And the best data enrichment services are ones that are built natively into the data mastering process.

Tamr Enrich is a fully-managed service within Tamr Mastering that enhances existing datasets with additional, often external, data sources. Unlike traditional data marketplaces, Tamr Enrichment simplifies the deployment process.

Powered by Tamr Mastering, the data mastering engine matches the enrichment attributes with internal records automatically and with high accuracy, even without a primary key.

As a result, companies can seamlessly embed trusted, high-quality external data insights into their data mastering pipeline, enabling better, more confident decision making and richer business value.

To learn more or to schedule a demo, please visit tamr.com.



Tamr recognized in Gartner’s Market Guide for Data Validation and Enrichment for Party Data

The research provides insights into data validation and enrichment services trends and best practices.

Learn more: [Gartner Research](#)

Gartner

Market Guide for Data Validation and Enrichment Services for Party Data

Published 28 July 2022 - ID G00736508 - 15 min read

By Analyst(s): Melody Chien

Initiatives: [Data Management Solutions](#)

Digital business requires accurate data at points of entry and at all times. Data validation and enrichment services are a critical investment to ensure data quality. Data and analytics leaders should leverage these services to enhance their data quality capabilities, especially for customer data.

Gartner

Table 1: Representative Vendors in the Data Validation and Enrichment Services Market
(Enlarged table in Appendix)

Vendors ↓	Headquarters ↓	Data Validation and Enrichment ↓
Bizmindz Technologies	India	Address, Email, Phone
Byteplant	Germany	Address, Email, Phone
Clearbit	U.S.	Address, Email, Phone, Social Profile, Business Entity
Data8	U.K.	Address, Email, Phone, Business Entity
Datactics	U.K.	Business Entity
Dun & Bradstreet	U.S.	Address, Email, Phone, Business Entity
Egon	Italy	Address, Email, Phone, Business Entity
Experian	U.K.	Address, Email, Phone, Business Entity
GBG	U.K.	Address, Email, Phone
IBM	U.S.	Address
Informatica	U.S.	Address, Email, Phone, Social Profile
Innovative Systems	U.S.	Address, Email, Phone, Social Profile, Business Entity
Lake B2B	U.S.	Address, Email, Phone, Social Profile, Business Entity
LexisNexis Risk Solutions	U.S.	Address, Email, Phone, Social Profile, Business Entity
Melissa	U.S.	Address, Email, Phone, Social Profile, Business Entity
MetricFox	Singapore	Address, Email, Phone, Social Profile
Precisely	U.S.	Address, Email, Phone, Business Entity
SAP	Germany	Address
Service Objects	U.S.	Address, Email, Phone, Business Entity
Smarty	U.S.	Address
Sun Tec India	India	Address, Email, Phone, Social Profile, Business Entity
Tamr	U.S.	Address, Email, Phone, Business Entity
TIBCO	U.S.	Address, Email, Phone
UPS	U.S.	Address
Validity	U.S.	Address, Email, Phone

Source: Gartner (July 2022)



Tamr is a highly scalable next-generation data mastering platform that allows customers to use data product templates to consolidate messy source data into clean, curated, analytics-ready datasets. Organizations benefit from Tamr Mastering, the industry's first suite of data product templates that combine human curation, patented machine learning, mastering rules and enrichment with first- and third-party data to accelerate business outcomes and deliver business-changing insights. Tamr's cloud-native and SaaS solutions enable industry leaders such as Toyota, Western Union and GSK get ahead and stay ahead in a rapidly changing competitor environment. Tamr's work in the public sector includes the U.S. Air Force, U.S. Army, Department of Homeland Security and various other federal and local government agencies.

Visit www.tamr.com and follow [@Tamr_Inc](https://twitter.com/Tamr_Inc) on Twitter and LinkedIn for more information about Tamr, its partners, and investors.